The Dissertation Committee for Yoori Kim Certifies that this is the approved version of the following Dissertation:

Single-Molecule Studies Reveal the Dynamics of DNA Repair and Transcription-associated Proteins

Committee:

Ilya J. Finkelstein, Supervisor

Rick Russell

Karen M. Vasquez

Kevin N. Dalby

Seongmin Lee

Single-Molecule Studies Reveal the Dynamics of DNA Repair and Transcription-associated Proteins

by

Yoori Kim

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2018

Dedication

I dedicate this dissertation to my wonderful family.

Acknowledgements

Most of all, I sincerely express my deepest gratitude to my wonderful advisor Dr. Ilya Finkelstein for giving me the great opportunity to work in his lab. His guidance and teaching truly inspired me to pursue a real scientist with a critical and objective view of science. I am thankful to him for assisting me in learning exciting projects, writing this thesis, and winning the HHMI fellowship award for an international graduate student.

I would like to thank my dissertation committee members: Dr. Rick Russell, Dr. Karen Vasquez, Dr. Kevin Dalby, and Dr. Seongmin Lee for providing me with their valuable comments and time over the years.

I would like to thank my former and current colleagues: Logan, Max, Nacho, Steve, and Michael for helpful discussions and instructions. I also thank Armando, Jeff, Kaylee, Jim for creating a fun environment for me to enjoy the graduate life in the lab. I also appreciate Dr. Soo-Hyun Yang and Dr. Jonghwan Kim for their thoughtful advice and encouragements.

Finally, I would like to give special thanks to my mother, Mihwa Baek for endless prayers and love for me, my father, Kisun Kim for his love, and my brother, Sejong Kim who is a delight in my life. I also thank my maternal grandmother, Yeong-sun Lee for her support and love in my life. With their unchanging trust and support, I could continue to keep positive thoughts throughout my graduate studies.

Single-Molecule Studies Reveal the Dynamics of DNA Repair and Transcription-associated Proteins

Yoori Kim, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Ilya J. Finkelstein

Intrinsically disordered regions (IDR) are protein segments that lack a defined tertiary structure. IDRs are enriched in eukaryotic chromatin-binding proteins, where they modulate protein-DNA and protein-protein interactions. In this thesis, I probe the function(s) of IDRs via two case studies: the yeast mismatch repair (MMR) protein Mlh1-Pms1 and transcription factors (TFs) derived from *C. albicans*, a pathogenic yeast. Using single-molecule DNA curtain assays, I demonstrate novel roles for IDRs in promoting facilitated diffusion of Mlh1-Pms1 on DNA. IDRs improve Mlh1-Pms1's ability to bypass a single nucleosome and to navigate dense nucleosome arrays that resemble chromatin. Moreover, these IDRs are critical for the Mlh1-Pms1 ATPase activity and also for nicking of the DNA substrate. I propose that conformational changes in the Mlh1-Pms1 IDRs alter DNA interactions and the nucleolytic activity of neighboring structured domains. I also examine the dynamics of PCNA, another essential MMR factor, in the context of trinucleotide repeat (TNR) instability. I show that Replication Factor C preferentially loads PCNA onto $(CAG)_{13}$ structures. The $(CAG)_{13}$ repeat captures the loaded PCNA and prevents PCNA from diffusing. Lastly, I reveal a novel role for IDRs in DNA condensation by studying Efg1, a TF that regulates a cell-type switching network in *C. albicans*. Efg1 encodes a specific IDR of low complexity, referred to as the prion-like domain (PrLD). I

show that the PrLD is critical for the DNA condensation and recruiting other PrLD-containing TFs, wherein nucleosomes regulate the TF-DNA dynamics. I propose a model where transcription factors become concentrated via phase separation and bring gene regulatory elements together to promote gene activation. Overall, this study provides mechanistic insights into the functions of IDRs in the dynamic behavior of DNA-binding proteins.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Intrinsically Disordered Regions in Chromatin-Binding Proteins

Significance of intrinsically disordered regions

Proteins encode both structured domains and **i**ntrinsically **d**isordered **r**egions (**IDR**) (Figure 1.1A). Structured domains undergo cooperative folding into a unique tertiary conformation. In contrast, IDRs are incapable of folding into a well-defined three-dimensional (3D) structure. This is because IDRs do not have an amino acid composition that can provide long-range hydrophobic interaction to mediate cooperative folding. Instead, IDRs are typically comprised of a higher portion of polar and charged residues or low sequence complexity (1). At least 30% of the eukaryotic proteome is comprised of IDRs (Figure 1.1B). In contrast to structured domains, however, it was not until recently that IDRs were recognized as a critical region for contributing to proteins' roles in cells.

IDRs can be found in any part of a protein, including carboxyl- and amino-terminal regions, or present as linkers connecting structured domains. The length of IDRs can span from ~10 to hundreds of amino acid residues. No matter what the position or length of an IDR is, it is useful to classify IDRs based on their biological functions. In this section, I will first describe established ideas about how IDRs contribute to protein function. I will then discuss chromatin-binding proteins by focusing on their IDRs and target search activity.

Functional roles of IDRs

IDRs are now recognized as a functional unit apart from structured domains (1, 2) (Figure 1.1C). Several proteome-wide datasets revealed that "hub" proteins that interact with multiple partners tend to encode IDRs (3, 4). Because IDRs are inherently flexible, they can adopt various conformations that fit many targets, allowing IDRs to mediate the functions of proteins orchestrating complex cellular signaling pathways (5–7). For example, calmodulin has a central disordered region that interacts with over thirty proteins in response to increased cellular Ca^{2+} levels, enabling calcium-based cellular responses such as glycogen breakdown and smooth muscle contraction (Figure 1.1A) (3, 8, 9). Also, the CREB-binding protein, a transcriptional coactivator, is comprised of long IDRs connecting globular domains and interacts with hundreds of partners, which allows the gene regulation of processes as diverse as embryonic development and hypoxic response (10–12). A central feature of these IDRs is the presence of molecular recognition features that enable binding to many partners, facilitating the regulatory complexity of eukaryotic systems (2, 13).

Recent studies have established that in eukaryotic cells, IDRs with a low-complexity sequence can form micron-scale compartments that lack surrounding membranes (14–16). These structures function to locally concentrate proteins and nucleic acids into so-called biomolecular condensates. The condensates are involved in diverse cellular processes, including ribosome biogenesis, DNA damage, meiotic cell division, and stress resistance (15, 17–20). The assembling principle for such compartment has been described as liquid-liquid phase separation by which biomolecules are separated into a dense phase from a diluted phase in solution (14, 15). These IDRs are enriched with multivalent elements that govern intra- or inter-molecular interactions, providing the

propensity to naturally assemble into liquid-like droplets in solution or membraneless organelles in cells (16, 21, 22).

IDRs also perform additional regulatory roles through multiple mechanisms. First, IDRs can modulate an activity of other parts within the same protein through allosteric modulation or competitive binding. For example, a C-terminal IDR of glutamate receptor illustrates that zinc binding to the IDR induces allosteric inhibition of channel gating (23). In the case of HMGB1, the IDR inhibits DNA binding activity by occluding the basic DNA binding surfaces, preventing it from regulating gene expression (24). Secondly, IDRs are frequently subjected to post-translational modifications (PTM) that changes the stability or functional states of a protein as seen in histone tails or in cytoplasmic tails of receptor tyrosine kinase (2, 25–27).

IDRs have also been identified in a variety of DNA-binding proteins (28–33). Here, IDRs engage in diverse pathways mostly by adjusting target binding or post-translational modifications. For example, the CtIP/Sae2 family of DNA double-strand break repair proteins has IDRs that are required for bridging two DNA strands and interacting with other key proteins in the pathway (28). IDRs in single-strand binding proteins regulate cooperative binding on ssDNA (34). IDRs in histone chaperones contribute to chromatin remodeling by regulating their dynamic binding to histones and re-positioning nucleosomes through post-translational modifications (29). Moreover, phosphorylation of IDRs in transcription factor Ets1 reduces the DNA binding affinity (35).

Although many studies have uncovered IDRs that modulate binding activity for DNA or target proteins, how such IDRs play a role in the dynamic behavior of DNA-

binding proteins on chromatin is poorly explored. In this thesis, I investigate how IDRs impact the dynamic behavior and enzymatic functions of DNA-binding proteins in the context of chromatin.

Chromatin

Chromatin is a complex of genetic material composed of DNA and proteins required for the compaction of the nearly two-meter long human genome into the ~10 μm-diameter nucleus. The fundamental packaging unit of chromatin is the nucleosome: 147 base-pairs (bps) of DNA wrapped around a histone octamer that is comprised of two copies of each of the core histones – H2A, H2B, H3, and H4. In higher eukaryotes, there is an additional histone called H1 that further condenses the histone-DNA complex. Wrapping DNA around histone octamers ultimately leads to a ~10,000-fold condensed DNA structure, which is referred to as chromatin. Nucleosomes were first visualized as spherical complexes connected by DNA filaments ("beads-on-a-string") via electron microscopy (36) (Figure 1.2A). Further studies highlighted the importance of nucleosome occupancy in determining cell fate, DNA replication and repair, global gene expression patterns, and tumorigenesis (37–40). Our understanding of how the regulation of such chromatin-binding proteins leads to such cellular defects is an active area of modern research.

Rationale for efficient target search of chromatin-binding proteins

DNA replication, DNA repair, and gene transcription all initiate at a specific sequence or region that must be targeted by proteins. During the target search, such a protein is faced with thermodynamic and kinetic challenges (41): The thermodynamic

challenge means that a protein should recognize and tightly bind the target site among the 10^6 - 10^9 bp non-specific sites with single bp precision (41). The kinetic challenge refers to finding a target in mere seconds amidst the crowded chromatin occupied by nucleosomes and other protein machinery. Because the cellular processes are controlled through a protein signaling pathway, incorrect or untimely binding on chromatin can disrupt functional DNA repair or gene expression. Thus, efficient target search by chromatin-binding protein is necessary to avoid serious problems in cells.

Facilitated one-dimensional diffusion of chromatin-binding proteins

Rapid search for a specific chromatin site by proteins can be achieved by two non-mutually-exclusive pathways: one-dimensional (1D) diffusion along a DNA (sliding), and three-dimensional (3D) diffusion (42). A protein that binds DNA through 3D diffusion can perform a round of 1D search along the DNA molecule and vice versa. Experimental evidence of 1D diffusion was first provided by measuring the kinetics of lac repressor-operator interaction using filter binding assays (43). Following studies have revealed that the rate of specific binding to a target site was significantly increased by lengthening nonspecific DNA surrounding the site (41, 42, 44, 45). An *in vivo* single-molecule study has revealed that a LacI repressor spends ~90% of the time as nonspecifically bound to and diffusing on DNA during the target search for its operator (46). Furthermore, single-molecule techniques have directly shown that numerous proteins' target search on DNA is facilitated by 1D diffusion on DNA (47–49). Examples include the T7 RNA polymerase, Type III restriction enzyme EcoPI, and mismatch repair proteins such as Msh2-6, Msh2-3, and Mlh1-Pms1 (49–52).

1D movement by proteins on DNA has been subdivided into two distinct mechanisms: 1D sliding and hopping (Figure 1.2B) (53). **1D sliding** is a diffusion mode that follows the helical patch of the DNA backbone without complete dissociation. This mode is useful for a protein that needs to recognize a subtle difference in the DNA backbone (50, 51). **Hopping** refers to a series of microscopic dissociation and association of proteins: a model where non-specifically bound proteins dissociate from the DNA and then rapidly reassociate with the same or nearby DNA site. Notably, hopping is advantageous in that it can i) move to another proximal DNA strand without undergoing 3D diffusion, and ii) enable the bypass of obstacles on chromatin as it diffuses on DNA.

If a protein loses its target search capability entirely or has it replaced by a different one, it may become non-functional (50, 54). For example, Msh2-6 and Msh2-3, the lesion sensor proteins in DNA mismatch repair, each possess a mispair-binding domain (MBD) that confers a different target search mechanism. When the MBD of Msh3 was replaced with that of Msh6, Msh2-3 lost its hopping mechanism as well as its nucleosome bypass activity, which would greatly reduce its ability to carry out the first step of mismatch repair (50).

DNA Mismatch Repair

Significance of DNA mismatch repair

Accurate and synchronized replication of DNA is required for proper cell division. Errors during DNA replication cause heritable mutations that may be deleterious for the fitness of progeny. However, the replication machinery introduces an incorrect nucleotide

every ~10^7 nucleotides per cell cycle (55). Therefore, the diploid human genome will accumulate ~600 replication errors during replication of its ~6 billion nucleotides (55). The spectrum of replication errors ranges from **s**ingle **b**ase-pair **m**ismatches (SBMs; the base pairing with non-complementary bases) and short **i**nsertion/**d**eletion **l**oops (IDLs; protruded non-complementary single-stranded DNA) (56, 57). Cells have evolved a post-replicative **m**is**m**atch **r**epair (**MMR**) system to correct these errors (Figure 1.3). The MMR system increases the replication accuracy by 100-fold (58).

MMR is evolutionarily conserved and proteins capable of executing MMR are found in all kingdoms of life. Defects in MMR factors cause hypermutations in *E. coli*. In humans, defects in MMR proteins are the main cause of Lynch syndrome, which increases the risk of colon cancer as well as endometrial, ovary, stomach, small intestine, brain, and skin cancers (59, 60). Thus, functional MMR activity is required for maintaining genome integrity and to prevent human diseases.

MMR is also implicated in **t**ri**n**ucleotide **r**epeat (**TNR**) instability. An increased number of the repeats causes many neurological disorders (61). For example, Huntington's disease is caused by the expansion of $(CAG)_n$ repeats in the *htt* gene (62, 63). These repeats can form extrahelical DNA structures (Figure 1.6) and these structures tend to expand in affected tissues presumably due to aberrant DNA repair (64). To date, MMR seems to both potentiate and attenuate TNR-mediated instability (65, 66). Studies using an *in vitro* cell extract system have shown that MMR proteins can repair the TNR-driven hairpins on non-replicating DNA (64, 67). On the other hand, genetic studies have revealed that deletions of key MMR proteins resulted in dramatic reduction of TNR expansions in mammalian

cells (68–71). Intensive research on the relationship between MMR proteins and TNR instability are needed to understand why the normal repeat sequences become pathogenically long and how both normal and expanded TNRs are handled by the MMR system.

Overview of MMR mechanism

All prokaryotic and eukaryotic MMR share similar biochemical steps. MutS and its eukaryotic homologs act as lesion sensors. MutS homologs recognize various types of replication errors by sensing non-B-form and bent DNA (72). Then, MMR-specific endonucleases, MutL homologs or MutH in *E. coli*, excise the lesion-containing DNA strand. This excision step should occur on the specific DNA strand that contains a lesion, or the error will be transmitted during next DNA replication. Strand discrimination is achieved by the presence of methylation at a specific sequence GATC in *E. coli*, and by the orientation of PCNA in eukaryotes (73, 74). Next, the lesion-containing DNA is degraded by exonucleases, and DNA is resynthesized by PolIII in *E. coli*. and Polδ or Polε in eukaryotes (Figure 1.3). In this thesis, I will focus on eukaryotic MMR.

Eukaryotes have two lesion sensors that form heterodimers, Msh2-Msh6 (MutSα) and Msh2-Msh3 (MutSβ). They have distinct mispair-binding domains that tend to target different subsets of lesions, but also share some target overlap. Msh2-6 recognizes single base-pair mismatches (SBM) and 1 nt short insertion/deletion loops (IDL) whereas Msh2-3 recognizes larger IDLs (up to 15 nt) and some SBMs (75, 76). Msh complexes form a sliding clamp with ATP and are able to diffuse on DNA, but each complex has different target search mechanisms (50). Whereas Msh2-6 uses 1D sliding, Msh2-3 uses a hopping

mechanism that also allowed the protein to bypass nucleosome barriers to efficiently target a lesion on chromatin (50). Once the lesions are recognized by Msh complexes, endonucleases are recruited to excise the lesion-containing strand.

Mlh1-Pms1 (Mlh1-Pms2 in humans) is an endonuclease in eukaryotic MMR (77). This heterodimer complex is recruited to an MMR site after lesion recognition by Msh2-6 (78). The endonuclease activity of Mlh1-Pms1 requires PCNA loading by RFC on DNA. Mlh1-Pms1 is dispensable when a nick is present 5' to a mismatch, (79) but the presence of Mlh1-Pms1 increases the rate of MMR (80). A recent single-molecule study has shown the 1D diffusion of Mlh1-Pms1 on DNA (81). How the diffusing activity of Mlh1-Pms1 is connected to its multiple nicking activity is explored further in this thesis.

ExoI is a 5' to 3' exonuclease and can degrade DNA from a DNA end or a nick (82). ExoI is recruited to repair sites via interactions with Msh proteins and Mlh1 (83). ExoI is not strictly required *in vivo* as ExoI-null cells are MMR-proficient with only 2-4 fold increase in mutation rate (79, 84). Genetic analyses in yeast have revealed the existence of ExoI-dependent and ExoI-independent subpathways in MMR (79). The majority of mutations affecting ExoI-independent MMR cause defects in efficient Mlh1-Pms1 endonuclease. The interaction of ExoI with Mlh1-Pms1 was required in the ExoI-dependent MMR. (85, 86). It is clear that Mlh1-Pms1 functions are important in both MMR pathways to degrade the lesion-containing DNA strand. To complete the repair, DNA polymerase δ resynthesizes the DNA gaps made during nucleolytic degradation by Mlh1-Pms1 and ExoI, and the resulting nicks are sealed by DNA ligase (79).

Structure and Function of Mlh1-Pms1

Mlh1-Pms1 forms a ring-like structure comprised of long linker arms connecting N- and C- terminal domains (Figure 1.4). The N-terminal domains (NTDs) of Mlh1-Pms1 are highly conserved in sequence and possess ATP-binding capabilities (87, 88). Mutations in the NTD of human Mlh1-Pms2 has been associated with cancer predisposition and failed in mismatch correction *in vitro* (89–92). The ATPase of Mlh1-Pms1 has been classified into the GHKL family because the ATPase domains of DNA **G**yrase, **H**sp90, histidine **K**inase, and Mut**L** display similar features that are distinct from canonical ATP-binding folds such as Walker-type ATPases (88, 93). These proteins exist as dimers, and the ATPase domains self-associate upon ATP binding, forming an ATP-induced central cavity which helps to capture a DNA substrate (Figure 1.5). ATP hydrolysis does not directly power this ATPase domain dimerization (87, 88, 93). Studies using prokaryotic MutL have suggested that ATP induces local structural rearrangement around the ATP binding site, which results in NTD dimerization and causing the CTD and NTD to be in close proximity (87, 94, 95). Using limited proteolysis and yeast two-hybrid assays, P. Tran et. al. have shown that the NTD of yeast Mlh1, but not the CTD, undergoes a conformational change triggered by ATP, making NTD of Mlh1-Pms1 heterodimerize (92). Similar properties of ATP-induced conformational change have been observed in hMutLα (91, 96). AFM imaging directly showed that yeast Mlh1-Pms1 is mostly present in an 'open' extended form in the apo (no nucleotide) state, and transitions to the 'closed' form upon nucleotide binding (97). Aside from the "gate" formed by ATP-induced dimerization for DNA capture, the positive charges on the NTD are also thought to provide a groove for DNA binding via electrostatic interactions (98–100). Mlh1-Pms1 can bind both ds- and ss-DNA with higher

affinity for dsDNA, and its cooperative binding to DNA has been observed using AFM, marked by long track formation of the protein on dsDNA (98, 100, 101).

The CTD of Pms1 in yeast (Pms2 in human) has a conserved DQHA(X)2E(X4)E endonuclease motif (77, 102, 103). Point mutations within the motif of yeast Mlh1-Pms1 greatly increased mutation rates *in vivo*, and a missense mutation in the motif of hPms2 has been found in a family with Turcot syndrome (77, 104). Several studies have shown that *B. subtilis* MutL, *A. aeolicus* MutL (*aq*MutL), yeast and human MutLα, but not *E. coli* MutL, are Mn^{2+}–dependent endonucleases that are stimulated by ATP (77, 102, 103, 105, 106). The CTD from *aq*MutL and yeast Mlh1-Pms1 have shown very weak endonuclease activity compared to a full-length protein containing the NTD (103, 105, 106). However, the endonuclease-dead yeast Mlh1-Pms1 mutant did not affect ATPase activity or formation of a complex with mismatch-bound yeast Msh2-6 (77). These observations suggest that conformational rearrangements of Mlh1-Pms1 must occur in the presence of both NTD ATPase and the CTD to facilitate nicking activity (77, 102, 103).

Mlh1-Pms1 has linkers between the ATPase and dimerization domains. The linker arms seem to lack secondary structure based on circular dichroism spectroscopy and AFM imaging, which is consistent with the secondary structure prediction using PONDR a web-based predictor of IDRs (107). These linkers share no sequence similarity among MutL homologs (97, 108) (Figure 1.4). *E. coli* MutL has a proline-rich 100-residue linker that tolerates sequence substitutions; deleting up to one-third of its length did not reduce MMR activity (108). PONDR predicts that *S. cerevisiae* Mlh1 and Pms1 have 160-residue and 290-residue linkers, respectively (Figure 3.1). In addition to larger linker arms than MutL,

Mlh1-Pms1 possesses a nucleolytic function that is not present in *E. coli* MutL. Consistent with these observations, deletions of only one-sixth of each linker compromised MMR in *S. cerevisiae* (109). However, why the disordered linkers of Mlh1-Pms1 are essential for functional MMR *in vivo* is not known. Chapter 3 of this thesis investigates the mechanistic roles of the disordered linkers in Mlh1-Pms1.

Intrinsically Disordered Regions in Transcription

IDRs in transcriptional regulators

Computational analysis of datasets for transcription factors (TF) retrieved from Swiss-Prot indicates that > 90% of these TFs possess IDRs (110). This tendency was more prominent in eukaryotic TFs than in prokaryotes, reflecting the heightened complexity in eukaryotic gene regulation (110). Several studies have established that disordered regions of transcriptional coactivators CREB-binding proteins and p300 take up nearly 60% of the sequence and are indispensable for their ability to interact with hundreds of TFs to regulate numerous cellular signaling pathways (12, 111). IDRs in other TFs including ETS factors are important for tuning DNA binding affinity, allowing for precise regulation of ETS-mediated transcriptional activity (31, 32). Furthermore, a recent study has claimed that coactivators BRD4 and MED1 are concentrated through IDRs and forms a transcription apparatus at super-enhancer regions *in vivo* (112).

Gene controls in cell identity and disease

Enhancers are DNA regions that are bound by transcription factors (TFs) to activate gene expression and are approximately one to several hundred base-pairs in length (113).

Super-**e**nhancers (**SE**) are clusters of enhancers that drive the robust expression of genes controlling cell fate specification during various cellular processes including embryonic development (114–116). Enhancers in SEs possess unusually high occupancy of enhancer-associated factors such as Mediator and p300. They also have higher levels of acetylation of histone H3 lysine 27 and other enhancer-associated histone marks than regular enhancers. Since SEs were initially described in murine cell identity genes (114), it has been suggested that SEs play key roles in human cell identity and diseases by detecting histone H3K27ac modification as a SE marker (116). Indeed, most SEs in humans activate cell type-specific genes, and the vast majority of disease-associated SNPs (such as those involved in diabetes and Alzheimer's disease) and oncogenes are associated with SEs (116). However, no previous group has shown whether a human pathogen uses SEs to maximize virulence, transmissibility, or other aspects of pathogenicity.

A case study of IDR-containing TFs relevant to cell identity in a human pathogen

Candida albicans is a pathogenic yeast that is present in the human intestinal tract and mouth. It is normally commensal and harmless as a major member of human microbiome but can be infectious in severely immunocompromised patients (117). This fungal pathogen can reversibly switch between two cell types named "white" and "opaque", and each type is heritable (118). These two cell types display distinct properties in metabolic preferences and mating competence as well as cell structures despite their identical genetic sequences (119). This suggests that the cell type switching is regulated epigenetically. Several studies established that the white-opaque switching is controlled by a transcriptional network comprising at least six master TFs including Efg1, Wor1, Wor3,

Wor4, Czf1, and Ahr1 (120–122). I propose that the TF network in cell type switching in *C. albicans* likely co-occupies a SE for several reasons: i) SEs are known to control cell identity, ii) ChIP analysis showed that the target genes in the network are bound by multiple TFs even though each TF paradoxically has its own DNA binding motif (122), and iii) the TFs directly activate or repress other TFs in the network (forming both positive and negative feedback loops) in a way that is associated with the timing of white-opaque transition. Although the cell type is not directly correlated to pathogenicity, it is clear that the TF that is controlling the cell type switching is also responsible for the virulence of this species (123, 124).

Strikingly, the master TFs in the white-opaque circuit contain large disordered regions (See Chapter 5). I hypothesize that the disordered regions are the key identities by which the TFs in this cell type switching create a super-enhancer that is a concentrated transcriptional apparatus to regulate a set of genes for a cell type. In Chapter 5, I show and discuss how IDRs can play a role in this SE context (Figure 1.7).

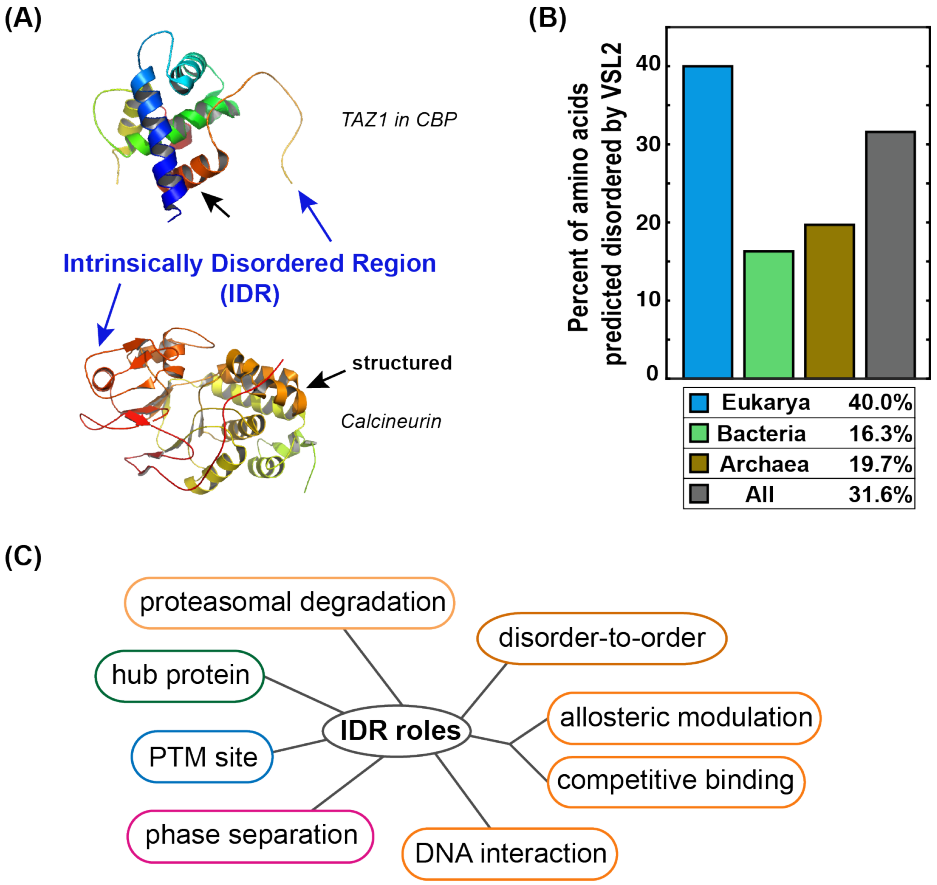

***Figure 1.1*: Overview of IDRs in proteins**. (A) Presentation of disordered and structured regions in two cases: a transcription coactivator CBP (10) and calcineurin (125). (PDB 2KA6 and PDB 5C1V). Representative IDRs and structured parts are indicated in blue and black arrows, respectively. (B) Predicted disordered residues by VSL2 in three domains of life (126). IDRs tend to increase with increasing complexity, comprising more than 30% of proteome in eukaryotes. (C) Schematic of various roles of IDRs in cellular processes.

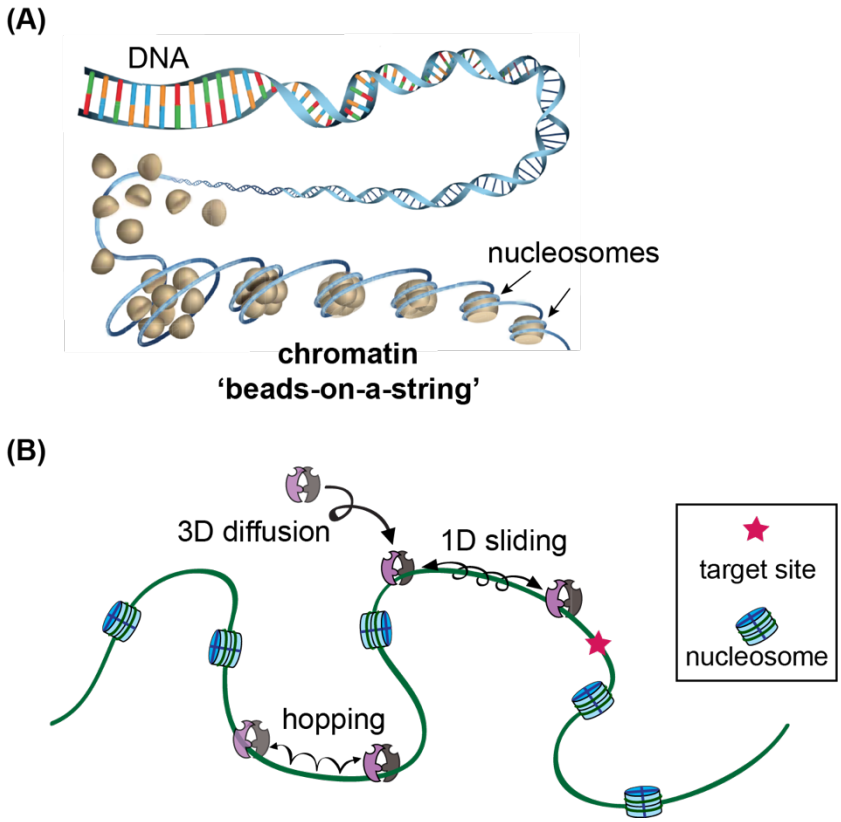

***Figure 1.2*: Chromatin and target search mechanisms of chromatin-binding proteins**. (A) Schematic representation of chromatin. DNA is assembled with nucleosomes, which forms 'beads-on-a-string'. Adapted from Ref. (127). (B) Schematic of target search mechanisms of proteins on chromatin. Many DNA-binding proteins can interact with chromatin at random sites via 3D diffusion from non-chromatin areas. Once a protein bind DNA, it can slide along the helical track of DNA (1D sliding), or it can move with microscopic dissociation and re-association from DNA (hopping) (128).

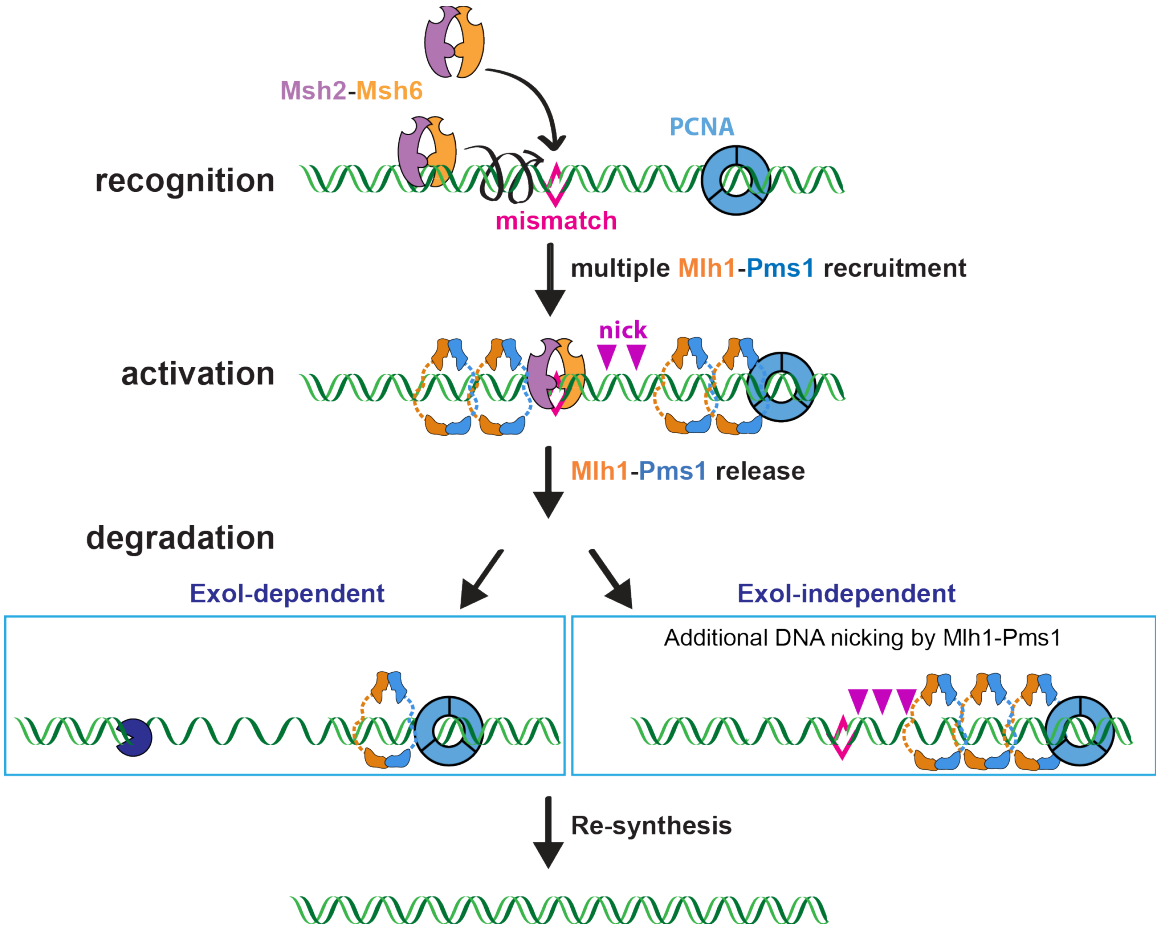

Figure 1.3: **Two models of eukaryotic DNA mismatch repair**. The major MMR proteins sequentially act on DNA region that misincorporated a mismatched base during DNA replication. First, Msh2-Msh6 recognize a mismatch through 3D and 1D diffusion in the vicinity of replication fork. Mismatch-bound Msh2-6 recruits multiple Mlh1-Pms1. The endonuclease Mlh1-Pms1 is activated by PCNA to produce nicks on DNA. PCNA also serve as a sensor for strand discrimination. In ExoI-dependent model, Exo1 enters through the nicks for further degradation on the lesion-containing strand. In ExoI-independent pathway, Mlh1-Pms1 stimulated by PCNA performs multiples rounds of nicking resulting in DNA degradation past the mispair (79). The gaps are filled by Polδ and remaining nicks are sealed by DNA ligase.

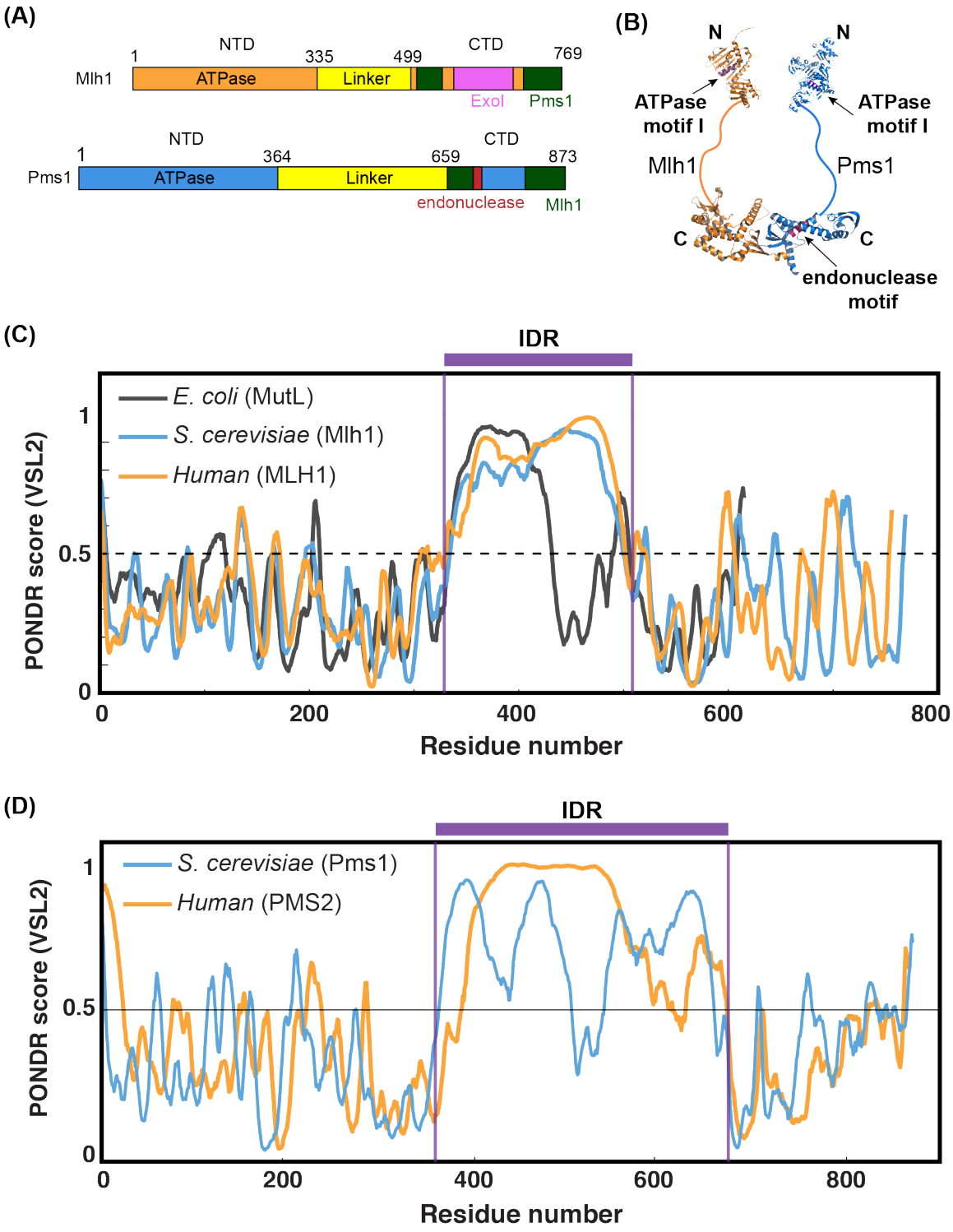
(A)
NTD
CTD
1
335
499
769
Mlh1
ATPase
Linker
Exol
Pms1
NTD
CTD
1
364
659
873
Pms1
ATPase
Linker
endonuclease
Mlh1
(B)
N
N
ATPase motif I
ATPase motif I
Mlh1
Pms1
C
C
endonuclease motif
(C)
IDR
E. coli (MutL)
S. cerevisiae (Mlh1)
Human (MLH1)
PONDR score (VSL2)
1
0.5
0
0
200
400
600
800
Residue number
(D)
IDR
S. cerevisiae (Pms1)
Human (PMS2)
PONDR score (VSL2)
1
0.5
0
0
200
400
600
800
Residue number

Figure 1.4:

***Figure 1.4*: Structure of Mlh1-Pms1 and prediction of IDR linkers in MutL homologs**. (A) Domain structure of yeast Mlh1-Pms1 (99, 103). N-terminal regions are ATPase in both subunits. C-terminal domains have dimerization interfaces for interacting each other (green). Mlh1 CTD has binding site for ExoI (pink), and Pms1 CTD has the endonuclease site (red). (B) Illustration of Mlh1-Pms1 structure based on structural and biochemical data (88, 99, 103). N-terminal domains have ATP binding motifs, C-terminal domain of Pms1 has an endonuclease motif. The N- and C- terminal domains are connected by long flexible linkers, and the two subunits dimerize through C-terminal domains. (C) IDR prediction analysis of the MutL (E. coli), Mlh1 (yeast and human), and (D) Pms1 (yeast) and PMS2 (human) using PONDR (107). All three species have long IDRs in the middle of the proteins.

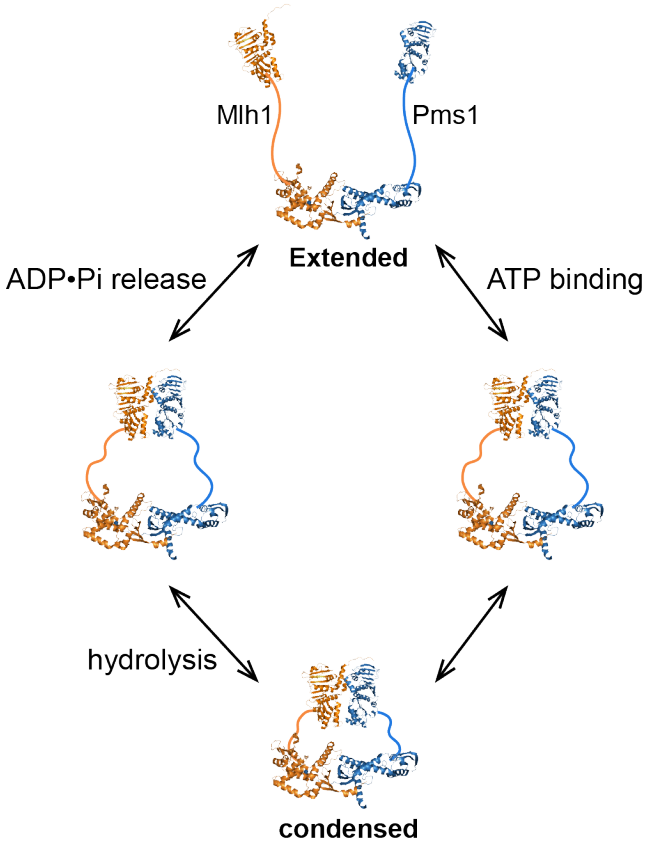

***Figure 1.5*: ATP-induced conformational change of Mlh1-Pms1.** Illustration of conformational change by ATPase cycle of Mlh1-Pms1 based on structural and biochemical data. In the absence of ATP, the complex is present as open, extended form. ATP binding to Mlh1-Pms1 induces dimerization of N-terminal domains and compaction of the dimer until hydrolysis takes place (97, 129).

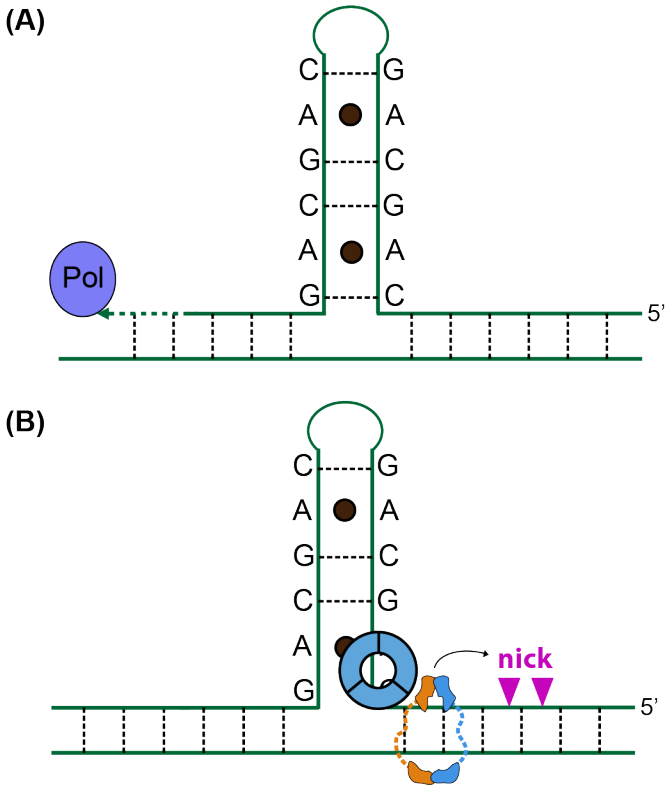

Figure 1.6: **Hypothetical model of trinucleotide repeat (TNR) instability**. (A) TNR instability caused by DNA polymerase slippage (61). (B) Hypothetical model of PCNA loading on TNR and thus MutLα–dependent nicking (130). As well as ssDNA/dsDNA junction, RFC can load PCNA onto non-complementary nucleotides such as TNR-driven stem-loop structures, which can recruit Mlh1-Pms1 to introduce nicks near the extrahelical structure (130).

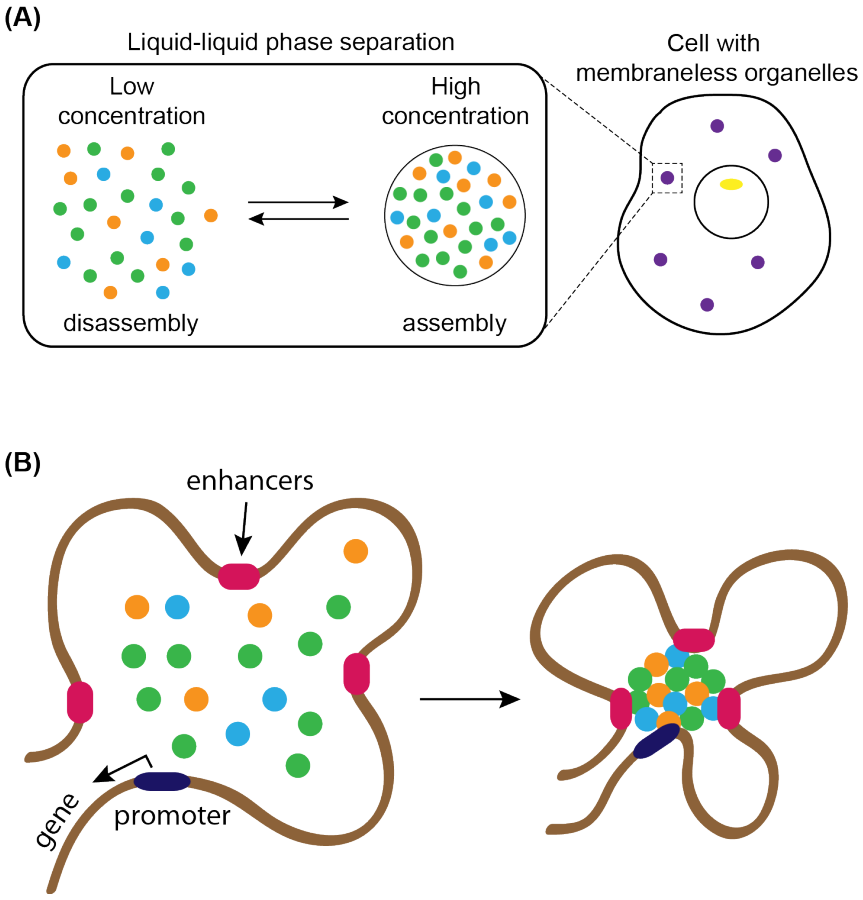

***Figure 1.7*: Phase-separation (LLPS) model for biomolecular condensates**. (A) Biomolecules with multivalency start condensing and phase separate in a concentration manner. The phase-separated condensates are found as liquid droplets *in vitro* and membraneless organelles in a cell. The components within the membraneless organelles are dynamic and can be exchangeable with exterior environment. (B) Schematic representation of the phase-separation model for transcriptional control. Transcription factors and mediators can form the multi-molecular complex of transcriptional regulators at a super-enhancer region *(115)*.

Chapter 2. Reconstitution of Human Nucleosomes on DNA Curtains

Parts of this chapter have been published in the following journal:

Brown,M.W., Kim,Y., Williams,G.M., Huck,J.D., Surtees,J.A. and Finkelstein,I.J. (2016) Dynamic DNA binding licenses a repair factor to bypass roadblocks in search of DNA lesions. Nature Communications., 7, 10607

– *Y.K. purified human histones, reconstituted nucleosomes, and characterized nucleosomes in ensemble and single-molecule assays.*

Introduction

Because DNA mostly exists as chromatin in eukaryotes, it is important to study DNA-related processes in the context of nucleosomes as well as on naked DNA. Therefore, several ways to reconstruct nucleosomal DNA substrates have been developed (131–133). Using these reconstituted nucleosomes for *in vitro* studies is useful because (i) they can be prepared with well-defined components of purified histones and DNA, and (ii) binding of proteins to nucleosomes or distinct species of nucleosomes can be easily detected by a non-denaturing gel-shift assay. There are two methods that have been widely used for nucleosome reconstitution: ATP-dependent assembly, which uses ATP-fueled enzymes, and ATP-independent assembly via step-wise salt dialysis (132–134). Both approaches require purified core histones and DNA to be assembled, but additional factors, ATP-utilizing protein CAF-1 and histone chaperon Nap-1 are required for the ATP-dependent assembly (131). Although enzyme-dependent nucleosome assembly is a good way to obtain periodic nucleosome arrays that are closer to physiological conditions, the salt dialysis approach has several advantages: (i) it requires only core histones and DNA, and (ii) the density of nucleosomes on DNA can be easily modified with changing

DNA:octamer ratio. The step dialysis method is achieved by gradually decreasing the salt concentration of a solution of the octamer and DNA from 2M to 0.2M NaCl (135). Protein interactions within a histone octamer are dominated by hydrophobic interaction and the histones form a stable octamer complex in high salt(136, 137). Upon lowering salt concentration from 2M to 1M NaCl, histone octamers start dissociating to $(H3\text{-}H4)_2$ tetramers and H2A-H2B dimers. At around 1M NaCl, tetramers stably bind DNA, whereas the dimers do not at this point. When the salt is further decreased to 0.6M NaCl, the dimers cap the tetramer-DNA complex to complete octamer formation(138). Since the gradient dialysis strategy form relatively irregular nucleosome arrays on DNA compared to the enzyme-dependent method, tandem repeats of nucleosome positioning sequence (NPS) have been used to obtain nucleosome arrays positioned at defined sequences on DNA. The commonly used NPS includes the sea urchin 5S rRNA gene that has been found in nature, or a synthetic sequence called "601" with high affinity for nucleosome formation(139, 140). This strategy facilitated the use of salt dialysis method for reconstitution of nucleosome arrays in studying DNA-binding proteins on nucleosomal DNA(37, 141).

How chromatin-binding proteins deal with the nucleosome obstacles and find their particular target sequences or binding partners is an important question to address. To answer this fundamental question, I developed a method to reconstitute fluorescent human nucleosomes on DNA and visualize these complexes using the DNA curtains assay. I characterized the reconstituted human nucleosomes via biochemical and single-molecule DNA curtains assays. I also confirmed that the DNA binding distribution of human

nucleosomes on a 48kb DNA substrate is statistically indistinguishable from previous results of *Saccharomyces cerevisiae* nucleosomes.

Materials and Methods

Purification of wild type and recombinant hH2A

The wild type or 3xFLAG H2A plasmid was transformed into BL21(DE3) codon plus RIL cells (Agilent). A colony was inoculated into 50 mL LB with 50 μg mL^{-1} carbenicillin and 34 μg mL^{-1} chloramphenicol, and grown at 37°C overnight. Fifteen milliliters of the overnight culture were seeded into 1.5 L LB and grown in the presence of both antibiotics. When the culture reached an OD600 of 0.6, 0.2 mM IPTG was added and the induction continued at 37°C for 3.5 hours. Cells were harvested at 5000 g for 15 minutes, and resuspended in 150 mL lysis buffer (100 mM NaPO4 pH 8.0, 8 M urea, 10 mM DTT, 15 mg benzamidine). Urea was deionized (501-X8 resin, Bio-Rad) immediately before use. Cells were lysed by sonication on ice, and centrifuged at 12°C and 100,000 g for 30 minutes. A 100 mL column was packed with 25 mL of SP-Sepharose Fast Flow resin (GE Healthcare), washed with 10 column volumes (CV) of water, and equilibrated with 10 CV of wash buffer (100 mM NaPO4 pH 8.0, 7 M urea, 10 mM DTT, 0.3mM benzamidine). The 150 mL supernatant was added to the column and rotated for 1 hour at RT. The supernatant was washed with 5 CV of wash buffer, and eluted with five fractions of 5 mL elution buffer (100 mM $NaPO_4$, pH 8.0, 1M NaCl, 7 M urea, 10 mM DTT, 4mM benzamidine). The resulting 25 mL eluent was loaded onto a Superdex-200 column (GE Healthcare) equilibrated in SAU-100 buffer (20 mM NaAcetate pH 5.2, 7 M urea, 100 mM

NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol). Gel filtration was performed using 120 mL SAU-100 buffer and the histone-containing fractions were loaded onto a tandem Q/SP column (10mL each). After loading the histones, the tandem column was washed with 5 CV of SAU-100. The Q column was removed and 3xFLAG H2A was eluted with a gradient from 0% to 100% SAU-600 (20 mM NaAcetate pH 5.2, 7 M urea, 600 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol) over 20 CV. The eluate was fractionated in 1.2 mL fractions and the histone-containing fractions were confirmed by SDS-PAGE. The protein concentration was determined by running SDS-PAGE gel with BSA standards (Pierce Biotechnologies). Purified protein was lyophilized and stored in -80°C. Both wild type and 3xFLAG H2A proteins purified with similar elution profiles and final yields.

Purification of H2B, H3, H4

Each of the three histones was purified from inclusion bodies as previously described, with minor modifications(132). Briefly, each histone was overexpressed in BL21(DE3) codon plus RIL cells. Cells were grown at 37°C and 0.2 mM IPTG was added at OD600 = 0.6, followed by additional 3 hours of induction at 37°C. Cells were harvested by centrifugation at 5000 g for 20 min at room temperature. Cell pellets were suspended in 25 mL TW buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA), and stored at -80°C until use. Each pellet was thawed and diluted up to 35 mL total volume using TW2 buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol, 1 mM benzamidine, and 1% (w/v) Triton X-100). Cells were lysed by sonication on ice for 2 min (10s on – 50s off). To harvest the inclusion bodies, the lysate was centrifuged at 20,000 g for 20 minutes at 4°C. The pellets were rinsed with TW2 buffer by resuspending,

and centrifuged at 20,000g for 20 minutes at 4°C. The pellet was washed in same way twice using TW2 without Triton X-100, and the final pellet was stored at -80°C.

Each inclusion body pellet was mixed with 200 μL DMSO and 6.5 mL unfolding buffer (20 mM Tris–HCl, pH 7.5, 7 M guanidinium-HCl, and 10 mM DTT) by gently agitating for 1 hour at RT, and centrifuged at 20,000 g for 20 minutes at 4°C. This was repeated two more times, and the supernatant from each centrifugation was dialyzed against 1L urea buffer (10 mM Tris–HCl, pH 8.0, 7 M urea, 1 mM EDTA, 5 mM β-mercaptoethanol, and 100 mM NaCl for H2B, 200 mM NaCl for H3 and H4) using 3,500 or 7,000 MWCO dialysis tubing (SnakeSkin, Pierce Biotechnologies). A tandem Q/SP column was equilibrated with 10% buffer B (10 mM Tris–HCl, pH 8.0, 7 M urea, 1 mM EDTA, 1 mM DTT, 1 M NaCl; buffer A is same components without 1 M NaCl), and the dialyzed histones were loaded and washed with 10% buffer B. H2B was eluted from 10% to 40% buffer B over 20 CV in 200 minutes and H3 and H4 were eluted from 20% to 50% buffer B over 20CV. The purified histones were checked by SDS-PAGE, lyophilized, and stored at -20°C until use.

Histone octamer assembly

Each of the four histones was dissolved in unfolding buffer (20 mM Tris–HCl pH 7.5, 7 M guanidinium-HCl, and 10 mM DTT), and gently agitated for 1 hour at RT. Each histone was mixed in equimolar ratios (used 10% higher molar ratio for H2A/H2B relative to H3/H4), and adjusted to a final concentration of 1mg/mL. The mixture was dialyzed against refolding buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 5mM β-mercaptoethanol, 2M NaCl) using 3,500 MWCO dialysis tubing with several buffer exchanges over 48 hours.

The dialyzed mixture was centrifuged to remove aggregates, and concentrated using spin-concentrator (Amicon Ultra-15; Millipore) to make a final volume of about 1mL. Gel filtration over a (Superdex-200, GE Healthcare) using SAU-200 was performed to resolve histone octamers from dimers and tetramers in the refolding buffer. The octamer peak fractions were combined, concentrated using a 10,000 MWCO spin-concentrator (Amicon Ultra-4, Millipore), and flash frozen using liquid N_2. The resulting histone octamers were stored in -80°C until use.

Total Internal Reflection Fluorescence (TIRF) Microscopy

The inverted type Nikon Eclipse Ti microscope was built with homemade aligning optic system for our single-molecule study. A 488nm 200mW diode-pumped solid-state laser (Coherent) was used for excitation light source, and the beam was aligned to target a quartz prism placed on top of a flowcell with a thin layer of immersion oil (Figure 2.1). The beam power at the face of the prism was typically ~10mW. Fluorescent images from each green and red channel were detected with two EM-CCD detectors at 5.0Hz using a 200ms exposure time. The data collected from each channel were saved as 8-bit TIFF files respectively using NIS-Elements software (Nikon). TIRFM images were collected using a 60x water immersion objective lens (Nikon, 1.2 NA Plan Apo).

Nanofabrication and flowcells

Diffusion barriers were constructed on quartz wafer by UV photolithography. The 100 mm wafer was cleaned with Piranha solution and rinsed with deionized water three times. The cleaned substrate was treated with the HMDS (hexamethyldisilizane) at 150°C

in a vacuum chamber in order to promote photoresist (PR) adhesion. On the primer-treated surface, the positive photoresist AZ5209E was used for spin-coat at 4000rpm for 2 minutes. Patterns were written by UV lithography with exposure time of 7.0 seconds using a Cr mask and Suss Microtec MA6/BA6 Contact Aligner. After patterning, the resist was developed using AZ-726-MIF developer, followed by rinsing with deionized water. The substrate was then cleaned with O_2 plasma for 1 minute using March Asher. A thin layer of chromium was deposited using a Cooke E-beam evaporator, followed by lift-off using acetone for 30 minutes with sonication. The patterned wafer was then rinsed with isopropanol and dried with N_2, and cut into six 55 mm x 22mm slides. To make a flowcell, inlet and outlet ports were made by drilling with a drill press (Servo) equipped with a 1.4 mm diamond-coated drill bit. The slides were cleaned by soaking in 2% (v/v) Hellmanex overnight, 1M NaOH, 100% EtOH in order, and stored in 120°C vacuum oven until use. The cleaned slide was assembled with a coverslip through double-sided tape (~100μm thick, 3M). The total volume of the sample chamber was ~15μL.

Lipid bilayer and DNA curtains

For the diffusible lipid bilayer surface, lipids were purchased from Avanti Polar Lipids and liposomes were prepared as described (142). In brief, a mixture of DOPC (1,2-dioleoyl-*sn*-glycero-phosphocholine), 0.5% biotinylated-DPPE (1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine-N-(cap biotinyl)), and 8% mPEG 2000-DOPE (1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-N- [methoxy(polyethylene glycol)-2000]) were used to passivate the chamber surface. mPEG prevents nonspecific adsorption of QDs, and biotin-lipid tethers individual DNA substrates. Excess liposomes were removed with

buffer containing 10 mM Tris-HCl (pH 8.0) and 100 mM NaCl. The sample chamber was then rinsed with buffer A (40 mM Tris-HCl (pH 8.0), 60mM NaCl, 1 mM DTT, 1 mM $MgCl_2$, and 0.2 mg ml^{-1} BSA) for 5 minutes. Streptavidin (0.02 mg ml^{-1}) in buffer A was injected into the sample chamber and incubated for 10 minutes. After rinsing with additional buffer A, λ-DNA (5~10pM) labeled at one end with biotin was injected into the chamber, incubated for 10 minutes, and unbound DNA was removed by flushing with buffer A. Application of flow aligned the DNA molecules along the diffusion barriers, pushed them into the barriers, and stretched the molecules along the barriers (Figure 2.2).

DNA substrate preparation and DNA staining

All experiments used lambda DNA (NEB). To immobilize the DNA on the lipid-bilayer surface, biotinylated oligonucleotides were purchased from IDT. They were hybridized to 12-nucleotide overhangs at the end of lambda DNA by slowly cooling with thermo cycler after warming to 70°C for 15min. The mixtures were ligated using T4 DNA ligase and filtered over Superdex-1000 column to eliminate excess oligomers and other reaction components. To visualize DNA backbones, 0.16nM YOYO-1 with oxygen scavengers was injected into the flowcell.

Results

Assembling human octamers and gel-based assay

To fluorescently label the histone octamers for single-molecule imaging, plasmids containing wild-type human histones were sub-cloned to include an additional 3xFLAG or

a 3xHA (Hemaglutinin) epitope tag. Since the N-terminal tails of histones protrude from the nucleosome, the epitope tags have been incorporated into N-terminus of each histone (143). Histones were over-expressed in BL21-CodonPlus(DE3)-RIL cells that complements eukaryotic tRNA genes lacking in *E.coli* for their high-level protein expression in the T7 expression system. The wild type and recombinant histones were successfully over-expressed in *E.coli* (Figure 2.3A). Histone overexpression in *E.coli* accumulates dense insoluble aggregates called inclusion bodies, which often results from partially folded intermediates *in vivo* (144). Each histone was purified in denaturing buffer solution containing 7M urea to help the insoluble histones dissolve from inclusion bodies. (132) Several strategies are developed to avoid the formation of inclusion bodies by changing culture conditions (e.g., decreasing temperature during protein induction). (145) In some cases including histones, however, the inclusion bodies can be useful to collect large amount of biologically active proteins by restoring through solubilizing in a strong denaturant such as 7M urea followed by dialysis against a refolding buffer solution (146, 147).

I followed previous protocol to assemble human histone octamers (132). The four histones were combined in denaturing buffer at an equimolar ratio for H3, H4 with 10% more for H2A and H2B (Figure 2.3B). It has been reported that H3 and H4 first form the tetramer $H3_2$-$H4_2$, and then two sets of H2A-H2B dimers form the octamer (148, 149). Thus, in addition to the canonical octamer, multiple histone complexes can form during the reconstitution reaction (149). For this reason, I have filtered the histone mixture over a gel filtration column to separate the complexes as well as aggregates. Adding excess H2A

and H2B prevents the formation of free $H3_2$-$H4_2$ tetramers or hexamers, which are hardly separated from octamers in the subsequent gel filtration step. In contrast, H2A-H2B dimers are easily separated from the octamers (Figure 2.3C) (147). To see whether the octamers have been formed and segregated properly, I performed a denaturing SDS-PAGE analysis for both wild-type octamers and 3xFlag octamers. The bands of each octamer contained each histone showing H2A (or 3xFlag H2A), H2B, H3, and H4 with a 1.0 : 0.9 : 1.2 : 1.1 stoichiometry as indicated by relative intensities of each band from SDS-PAGE analysis (Figure 2.3D).

I next reconstituted the human octamers into nucleosomes following established protocols (132, 147). I reconstituted the histones onto a synthetic 145 bp nucleosome positioning sequence labeled "601-DNA" after Widom et. al (140). The "601-DNA" was identified through several rounds of selection that comprised of salt dialysis reconstitution, sucrose gradient separation, and sequencing of the selected DNA from $5x10^{12}$ random sequences(140, 150). The affinity of the 601 DNA for nucleosome was determined by measuring free energy of 601 DNA relative to that of 5S rRNA gene sequence, which showed 2.9kcal/mol (140). Thus, the 601 acts as a strong positioning sequence where histone octamers selectively bind, which has been widely used for positioning histone octamers at precise DNA locations (151–153). Hence, I chose to use this DNA in testing the biochemical properties of our human histone octamers. The PCR-amplified DNA were concentrated via concentrator (Zymo Research) and dissolved in 2M TE buffer, then mixed with human octamers for stepwise salt dialysis (Figure 2.4A) (132). The reconstituted nucleosomes were analyzed via electrophoretic mobility shift assay (EMSA) using a 5%

TBE native gel (Figure 2.4B). Wild-type nucleosome showed a single up-shift band and did not exhibit any super-shift in the presence of α-Flag antibody (lanes 9 and 10), indicating that the α-Flag antibody does not bind any regions of the wild-type human octamer. The octamer containing 3xFlag H2A also formed nucleosomes (Figure 2.4B lane 3). Unlike wild-type nucleosomes, the 3xFlag tagged nucleosomes displayed two distinct bands. I hypothesize that the two bands may come from partial proteolytic degradation of the epitope tag. It has been reported that N-terminally inserted epitope tags (Flag, HA etc.) are susceptible to proteolysis (154). To further prove the hypothesis, mass-spectrometry analysis can be performed to detect which residues are eliminated.

Next, I performed super-shifts with reconstituted nucleosomes and α-Flag antibodies. Because each nucleosome has up to six Flag epitopes, I could observe the step-wise shifted bands of 3xFlag nucleosomes as I increased the antibody concentration (Figure 2.4B lanes 4-7). The upper band of the two species from 3xFlag nucleosomes first disappeared with α-Flag antibody presumably because it is more intact species that has all functioning epitope tags. To verify that the epitope tag is providing a unique target for the α-Flag antibody, I also tested α-HA antibody at the highest concentration I used in our gel-shift assays. There was no super-shift band for this condition, which looked similar to the no antibody condition (lanes 3 and 8). I also confirmed that the DNA is not recognized by the α-Flag antibody (lane 11). I conclude that 3xFlag-H2A containing histone octamers can form nucleosomes and that these nucleosomes can be specifically targeted by α-Flag antibodies.

Imaging of human nucleosomes on DNA curtains

I next reconstituted the human nucleosomes on λ-DNA for our single-molecule studies. The experimental procedure for making human nucleosomes on λ-DNA is shown in Figure 2.5A. Wild-type λ-DNA is functionalized at one end with biotin for the purpose of immobilization to streptavidin surface in the flowcell. After removing all unnecessary components using gel filtration column, the functionalized λ-DNA was concentrated via isopropanol precipitation, then dissolved in TE buffer containing 2M NaCl to be ready for the stepwise salt gradient dialysis step. Human nucleosomes were reconstituted on the DNA via step-wise salt gradient dialysis and the DNA was assembled into DNA curtains.

To fluorescently visualize the human nucleosomes, I conjugated anti-Flag antibodies with fluorescent quantum dots (Qdots). Quantum dots (Qdots) are nanometer-scale (approximately ~10nm in diameter (155)) semiconductor crystals. A core nanocrystal made of CdSe is coated with another semiconducting shell such as ZnS, which provides better photostability(156). Qdots are superior single-molecule dyes because they are highly stable against photobleaching, allowing single-molecule imaging for over twenty minutes in our imaging conditions (155, 156). Furthermore, Qdots have narrow emission spectra with broad excitation spectra, so we can excite two spectrally-distinct Qdots with a single excitation source (e.g., imaging green and red Qdots with a single 488nm laser). This enables visualizing of two different proteins simultaneously. Although Qdots have excellent optical properties for imaging single-molecules, there are also shortcomings such as blinking. The blinking is the process of random switching between ‘on’ and ‘off’ states of the Qdot emission. For example, we can observe intermittent blank frames in a movie of individual Qdot-protein. This can be problematic for a continuous tracking of protein of

interest because we can lose information of its movement that is analyzed based on the imaging of particle tracking. This problem can be partially alleviated by adding DTT (dithiothreitol) or BME (β-mercaptoethanol) to our buffers (157). For fluorescent labeling, I conjugated biotinylated α-Flag antibodies with streptavidin-conjugated Qdots. The surface covered with streptavidin was blocked by injecting a buffer solution containing free biotins prior to nucleosome labeling. The antibody and streptavidin were pre-incubated at equimolar ratio right before use, and then injected into the flowcell to bind and visualize the nucleosomes on the DNA curtain (Figure 2.5B). The irregularly deposited Qdot-nucleosomes were observed on single DNA molecules, as confirmed by the retraction of DNA molecules when flow stopped (bottom panel).

Verification of human nucleosomes on 13x601 DNA curtains

To verify that I have reconstituted authentic nucleosomes and not histone aggregates or incomplete histone complexes, I used λ-DNA that incorporated the 13x601 array at a specific region (Figure 2.6A). The nucleosome-containing 13x601 λ-DNA was allowed to make a DNA curtain in the flowcell and was visualized with the biotin-streptavidin-Qdot strategy. As expected, I could observe a clear line of Qdots positioned at around 35kb from starting point of the diffusion barriers (Figure 2.6B). To quantitatively determine the position of nucleosomes bound on λ-DNA, I determined the positions of 767 individual human nucleosomes by measuring the distance in kilobases (kb) from the starting point of diffusion barriers. The positioning data was compiled into a histogram to which a single Gaussian curve was fit according to equation (1):

$$y = f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad (1)$$

where $y = f(x|\mu,\sigma)$ is the probability density function of a Gaussian distribution at each of the values in x with mean μ and standard deviation σ. The center of the peak of the curve was located at 34.3kb with a standard deviation of ±2.16, which matches the actual region of the 13x601 array comprising 2,330 bp on λ-DNA (Figure 2.6).

Characterization of the human nucleosomes for single-molecule study

Next, I have characterized the numbers of nucleosomes on single λ-DNA molecules. First, I selected a range of the ratios that produce countable numbers of nucleosomes on DNA (Figure 2.7). Because Qdots fluoresce intermittently (blinking), I used ten movie frames taken over two seconds to detect all of the fluorescent particles. I also selected only the particles that were retracted by flow-off in order to eliminate the ones attached to surface or the Cr barriers. The number of nucleosomes on each DNA was easily countable using a DNA:octamer ratio of 1:75 (Figure 2.7A). Using that ratio, I tested whether the nucleosome binding on the DNA follows a Poisson distribution which can be used to determine the cooperativity of protein-nucleic acid binding. If a DNA binding protein binds DNA in a non-cooperative way, then the distribution of the number of bound protein among bulk DNA should obey a Poisson distribution (158). Thus, I asked whether the formation of human nucleosomes on λ-DNA displays a Poisson distribution under the low-density nucleosome condition. To test this, I made a histogram as a function of number of nucleosomes per DNA, and used it to fit Poisson distribution that has the mean and standard deviation of the data according to equation (2):

$$f(x|\lambda) = \frac{\lambda^x}{x!}\, e^{-\lambda} \qquad (2)$$

where $f(x|\lambda)$ is the probability of a Poisson random variable x that represents the number of nucleosome binding occurring in a given space (λ-DNA), 0, 1, 2, 3 … and λ indicates the mean number of nucleosome formation on λ-DNA. The result showed that the deposition of the nucleosome on DNA fairly well obeyed a Poisson distribution (Figure 2.7B), which is consistent with a previous observation that binding of histone octamers onto DNA is independent of other nucleosome formation(159, 160). In addition, I observed a gradual increase in the number of nucleosomes on a single DNA as a function of an increasing the DNA:octamer ratio (Figure 2.7C). Above the ratio of 1:75, the nucleosomes were so densely packed that it was difficult to count them.

Having determined that I can deposit nucleosomes on λ-DNA, I asked how our positioning distribution of human nucleosomes was affected by DNA sequence. D. Tillo *et al*. showed that GC content is the most influential factor of intrinsic nucleosome occupancy *in vitro* (161). High GC content obviously tends to lack poly-A stretches that are rigid and bent, thereby excluding nucleosome formation on the featured sequence. To analyze the pattern of positions of human nucleosomes deposited on the DNA curtain, I used the 1:75 of DNA:octamer condition where approximately 4~6 distinguishable nucleosomes are present per DNA. To eliminate the ambiguity of the orientation of the DNA, I consistently used the biotinylated oligomers that hybridize to only one side of the DNA. Also, I selected the field of views where the DNA molecules are perfectly aligned at the barriers, and only the Qdots overlaid on YOYO-stained DNA molecules.

To determine the exact position of Qdot-nucleosomes on λ-DNA, each x, y coordinate of individual Qdots on DNA was extracted by fitting to a Gaussian curve. The coordinate values were subtracted from the position of the barriers to determine the position of single Qdot-nucleosome on the DNA. The position data from 2899 particles were compiled to a histogram and compared with probability of base contents or yeast nucleosome data previously published (162). Next, I produced an A-T or G-C probability of λ- DNA, and compared it with our positioning data (Figure 2.8A, top and middle panels). The left half of the genome exhibits a high frequency of GC bases. The middle region of the DNA around 24 kb has an especially low frequency of GC content, and this region is highly matched to lower occupancy in both human and yeast nucleosomes. This is consistent with the previous findings that GC content is the major determinant for nucleosomes forming on DNA in vitro. I compared the positioning result with yeast data (Figure 2.8A, bottom panel). The result indicates that the positioning of human nucleosomes is significantly well-correlated with yeast data, showing a Pearson's r = 0.64 (Figure 2.8B). Given that our data reproducibility coefficient never exceeds 0.64, it is reasonable to argue that the observed positioning pattern of human nucleosomes is significantly matched with that of yeast nucleosome. I speculate that the intrinsic variability results from the salt dialysis reconstitution that is known as random assembly of nucleosomes on DNA, and nucleosome by itself does not possess strong preference for specific sequences (140, 152).

Discussion

The reconstituted human nucleosomes on DNA curtain for single-molecule studies will be useful for single-molecule studies and will answer numerous questions with respect to the ability of DNA repair enzymes to deal with nucleosome obstacles. For example, yeast Mlh1-Pms1 (MutLα) heterodimer has long unstructured linker arms and freely diffuse on both naked and nucleosomal DNA (109, 163). Given that MutLα mutants whose linker arms are disrupted via TEV protease cleavage showed severe defects in MMR in vivo, it can be speculated that the biological role of the linker arm of MutLα is primarily to freely navigate whole genome by passing over nucleosomes (109, 163). We can also investigate how other MutL homologs such as Mlh1-Mlh3 (MutLγ) manages the nucleosome obstacles differently, reflecting the fact that the two homologs act in a distinctive manner in such as characteristic of binding to DNA substrate or dependence of other interacting partners(164). In addition, there is strong evidence that nucleosomes are controlling the DNA resection machineries that generates single-stranded DNA for homologous recombination in double-stranded break (DSB) DNA repair (165, 166). In conclusion, the development of fluorescent nucleosomes for single-molecule studies will allow us to observe detailed molecular mechanisms of DNA repair proteins in the context of nucleosomes.

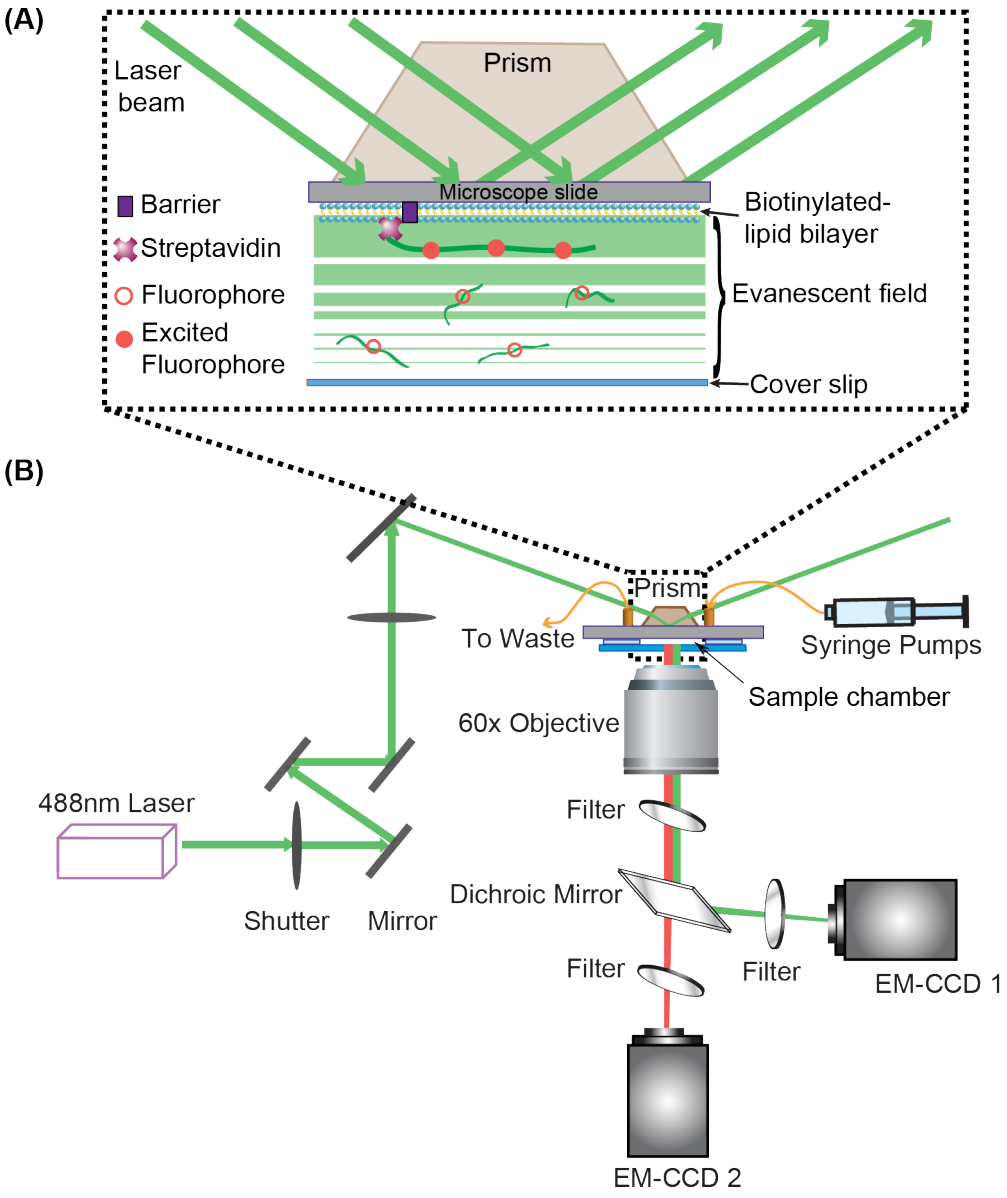

***Figure 2.1*: Microscope setup.** (A) Schematic of the imaging area using prism Total Internal Reflection Fluorescent (TIRF) microscopy. Incoming laser light and a prism are used to generate an evanescent field that selectively excites fluorescent molecules close to the surface. (B) Light path for prism TIRF Microscope. A 488nm laser was used for excitation light source and the beam paths were aligned to target the prism mounted on a flowcell. Two EM-CCD (Electron multiplying charge coupled device) cameras were installed to separately record two colors, green and red.

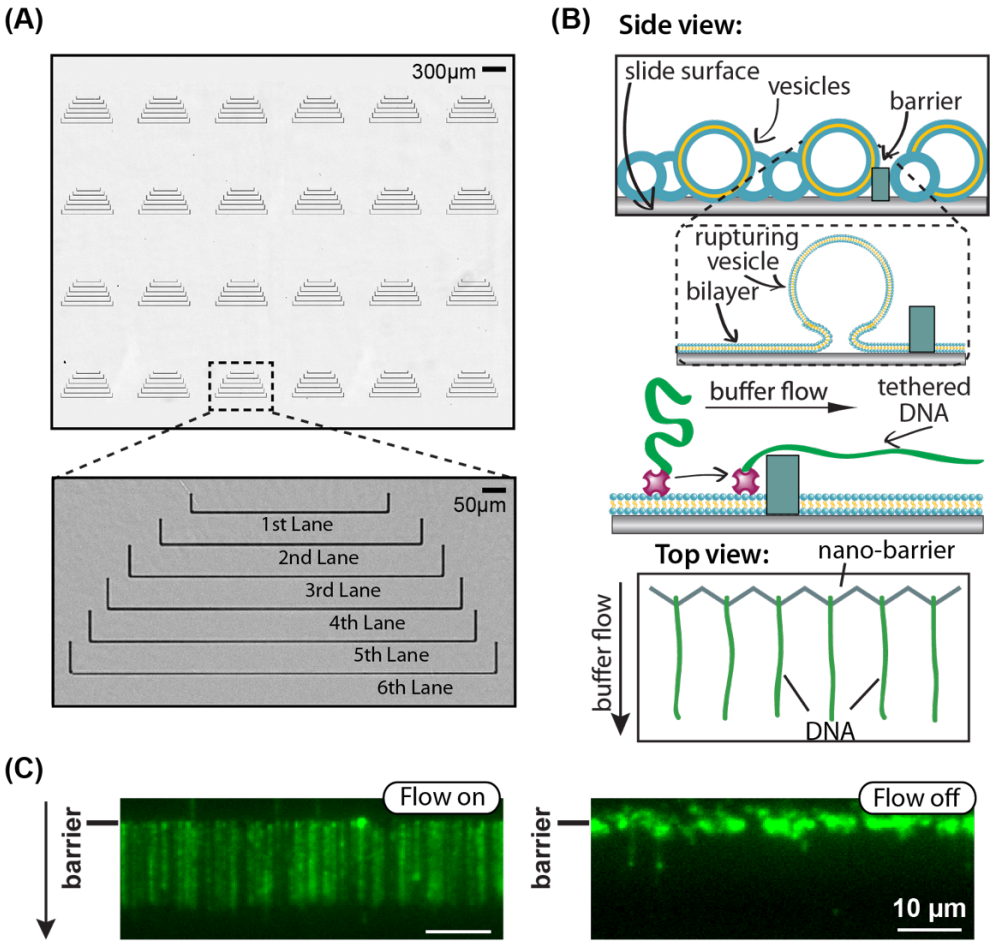

***Figure 2.2*: Experimental layout of making DNA curtain on a lipid bilayer flowcell**. (A) The images of the patterned chromium diffusion barriers. The top image shows an image of at 10x magnification of 6x4 series of barrier sets made of chromium deposited on a quartz slide. The bottom image shows a magnified image of a single barrier set with annotations. (B) Schematic of the principle of lipid bilayer that permits DNA curtain. The slide surface is covered by lipid bilayer containing 0.5% biotin-lipids, allowing streptavidin to mediate tethering of biotinylated λ-DNA to the lipid bilayer. As illustrated in the two bottom panels, the tethered DNA molecules are driven to the physical barriers when buffer flow is applied. (C) Fluorescent images of λ-DNA curtains aligned at the barriers. Individual λ-DNA molecules stained by YOYO-1 dye are shown in green. DNA is stretched during buffer flow (arrow in left image) and is retracted when buffer flow is blocked (right image).

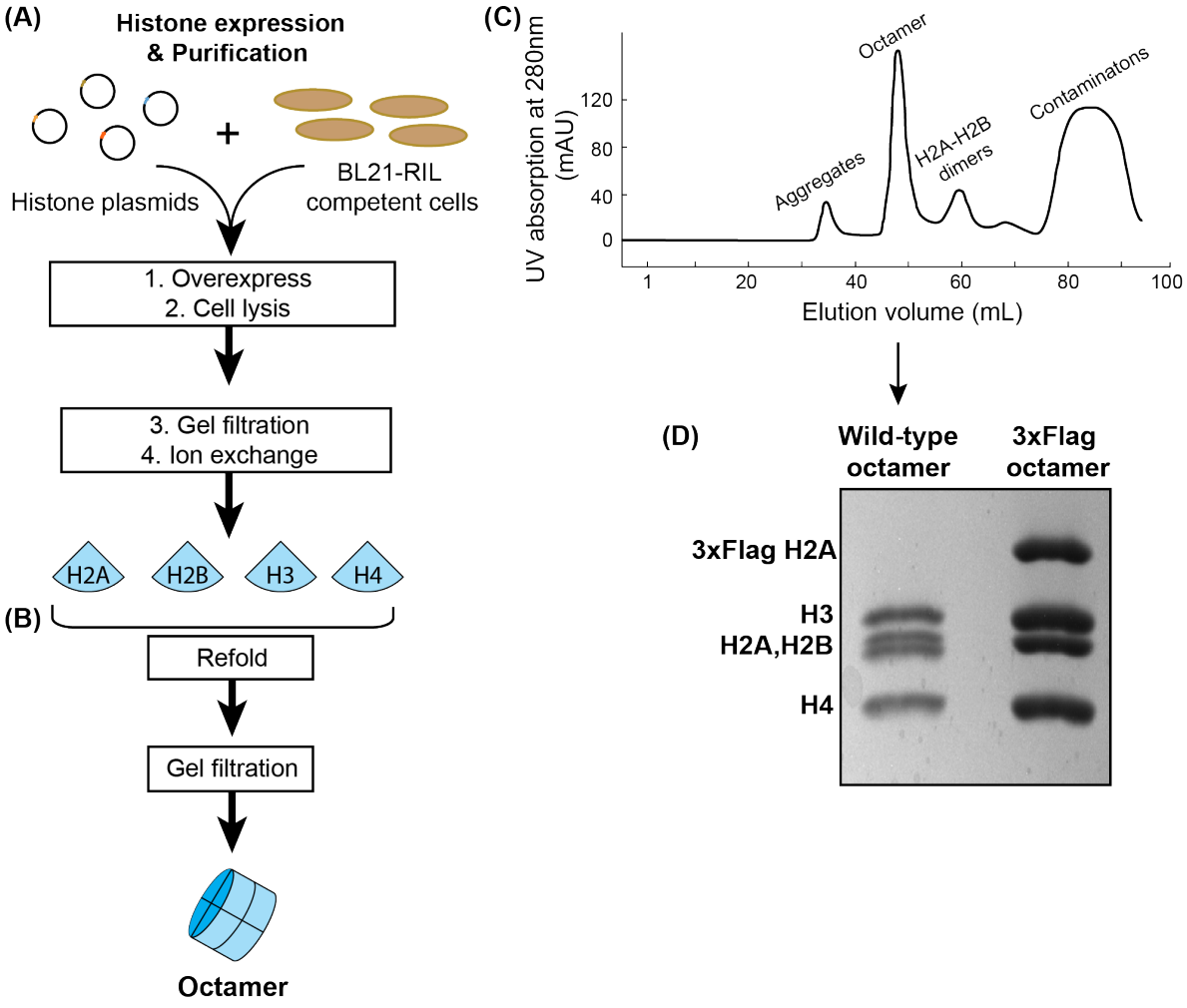

***Figure 2.3*: Human histone purification and octamer assembly**. (A) Schematic of the procedures for histone purification. Each wild-type or epitope-tagged histones was expressed in BL21-codon plus RIL cells and purified through gel filtration followed by ion exchange. (B) The four histones were mixed at equimolar ratio of H3 and H4 with 10% excess H2A and H2B to assemble histone octamers. The mixtures were filtered to eliminate aggregates and incompletely assembled octamers. (C) The elution profile of the gel filtration chromatography column. Fractions from the octamer peak were pooled and further analyzed. (D) A 15% SDS-PAGE analysis of wild-type and recombinant octamers used for reconstitution.

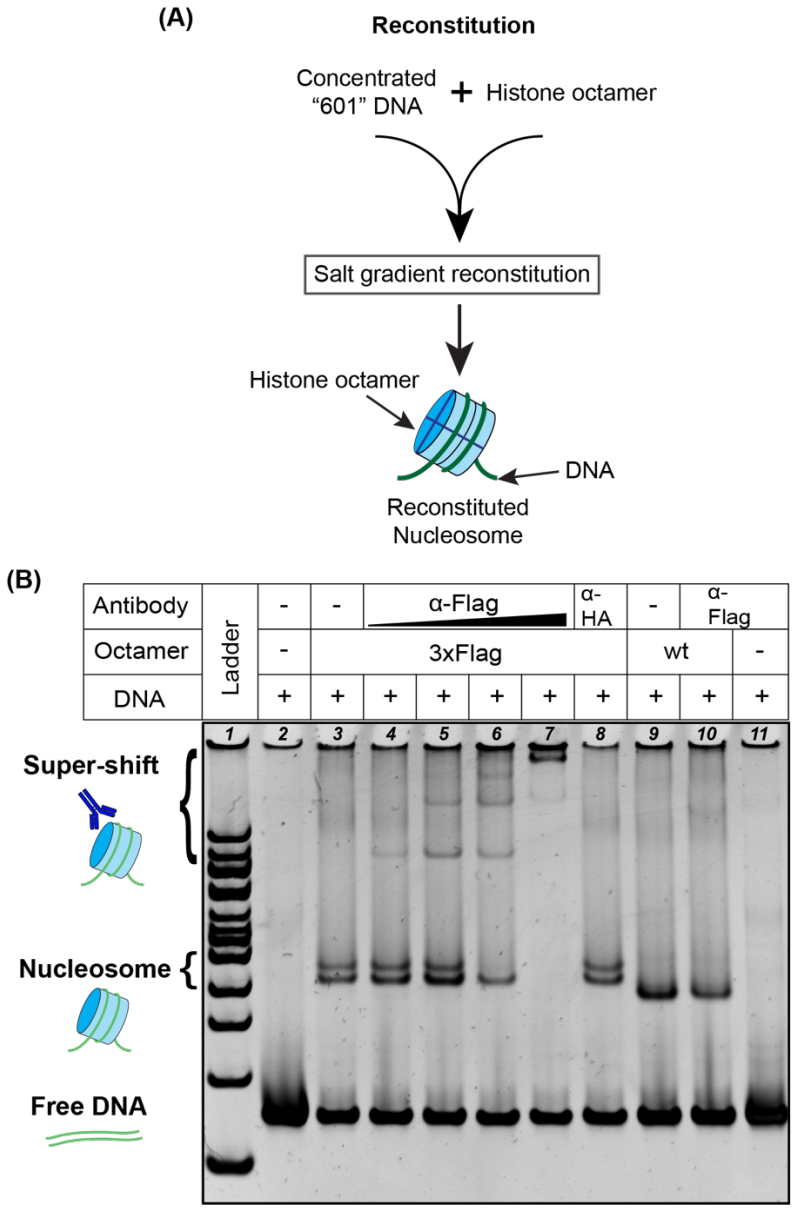

***Figure 2.4*: Human nucleosome reconstitution *in vitro* and native gel shift assays for antibody specificity**. (A) A high affinity DNA sequence for histone octamer called "601" was used for nucleosome reconstitution. The 145bp 601 DNA was mixed with the human histone octamers, and the human nucleosomes were prepared via step-wise salt dialysis. (B) 5% TBE native gel assay demonstrates that both wild type and 3xFlag tagged human nucleosomes were successfully reconstituted, and the α-Flag antibody specifically binds to the 3xFlag-tagged nucleosomes.

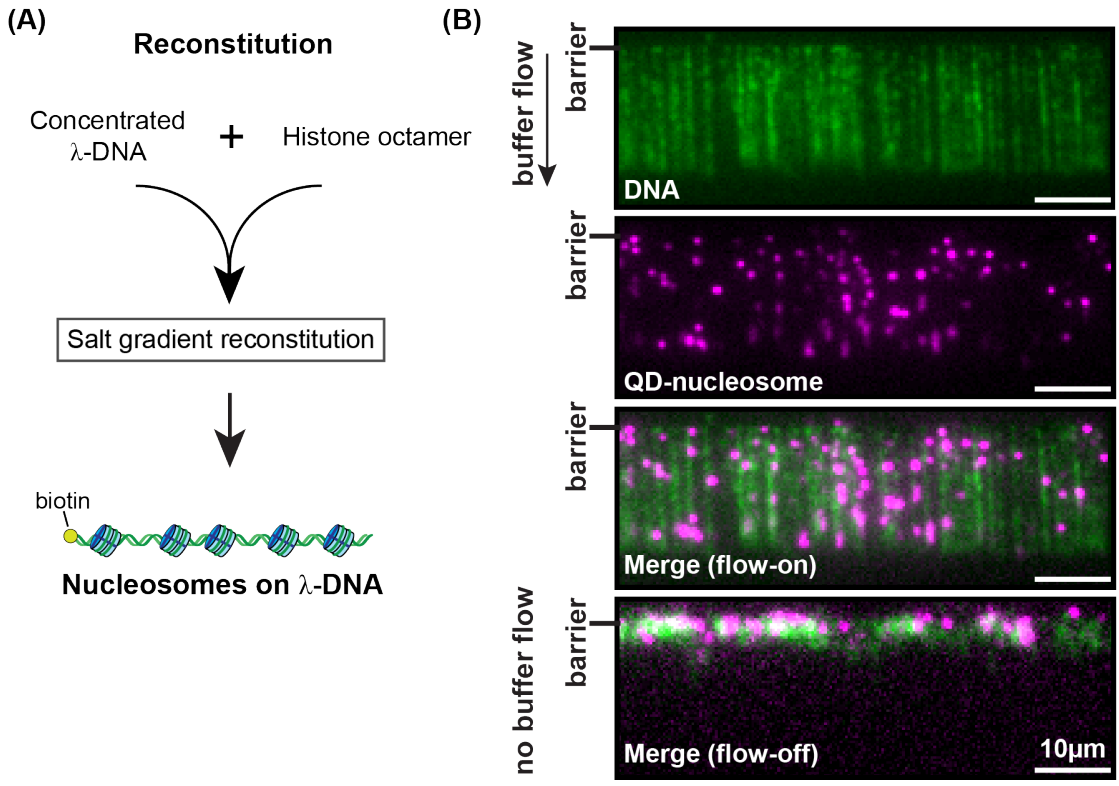

***Figure 2.5*: Visualizing human nucleosomes reconstituted on λ-DNA for single-molecule study**. (A) Schematics of the procedures to reconstitute the nucleosomes on λ-DNA. The λ-DNA was ligated with biotinylated oligomers at one end of the DNA to be tethered on flowcell surface. (B) Fluorescent images of human nucleosome deposited on λ-DNA. Nucleosomal DNA substrates are anchored to the flowcell surface and aligned along the nanofabricated barriers. YOYO-1 stained λ-DNA (green) is extended when buffer flow is applied (top panel). Individual human nucleosomes formed on the DNA are shown bound specifically to the DNA (magenta; second panel). All Qdot-nucleosomes were retracted as well as DNA when buffer flow stopped, indicating that the Qdots are on the nucleosomes formed on DNA, not on the surface (third and fourth panel).

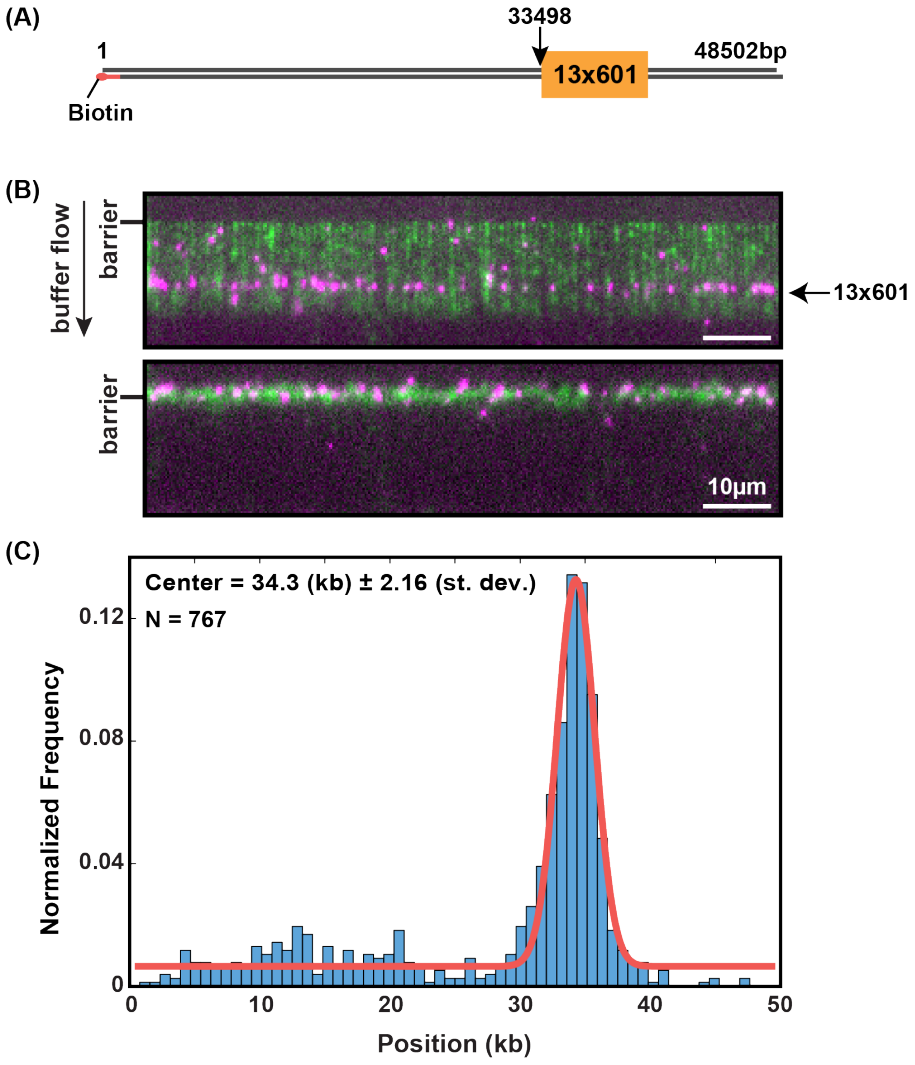

***Figure 2.6*: Verification of human nucleosomes reconstituted on 13x601 λ-DNA**. (A) Schematic of 13x601 λ-DNA substrates. (B) Image of DNA curtain of human nucleosomes that were specifically formed at 13x601 region on λ-DNA. The 13x601 sequence (2,330 bp) was incorporated via Xho1 and NheI restriction sites on λ-DNA. All nucleosomes were retracted when the buffer flow was turned off (bottom panel), indicating that all Qdot-nucleosomes deposited on 13x601 are on DNA. (C) Positioning distribution histogram of human nucleosomes deposited on 13x601 λ-DNA. The binding density of nucleosomes on DNA (blue) was fit to a Gaussian distribution (red) that shows the center of 34.3 (kb) ± 2.16, indicating that the nucleosome was assembled at the correct 13x601 position.

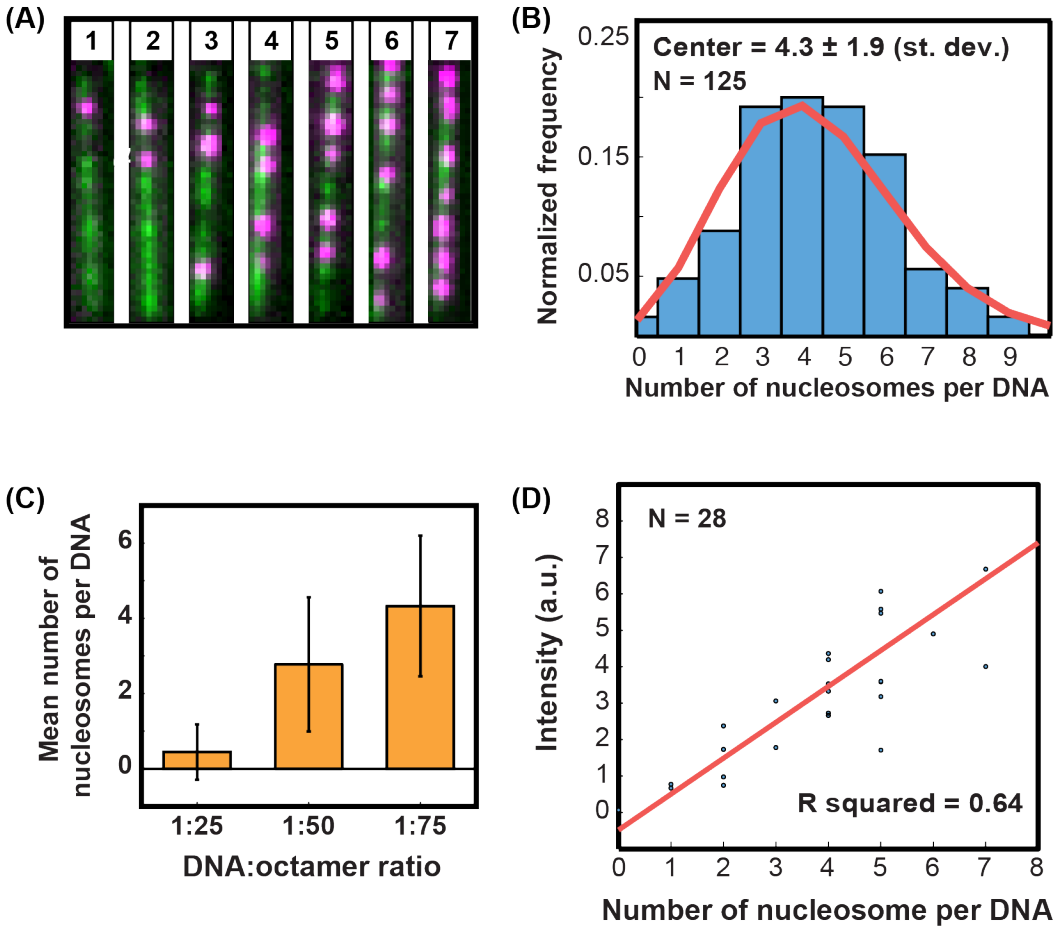

***Figure 2.7*: Counting of reconstituted human nucleosomes on wild-type λ-DNA**. (A) Representative images of various numbers of nucleosomes per DNA. (B) Histogram of frequency of nucleosome binding per DNA (blue) was fit to a Poisson distribution with the mean of the data (red). (C) The number of nucleosomes as a function of DNA:octamer ratio. The standard deviation of each data set is shown as a black line. (D) The intensity of Qdots per DNA as a function of number of nucleosomes per DNA. The measured intensities were plotted (blue circle) and fit to a linear model (red line).

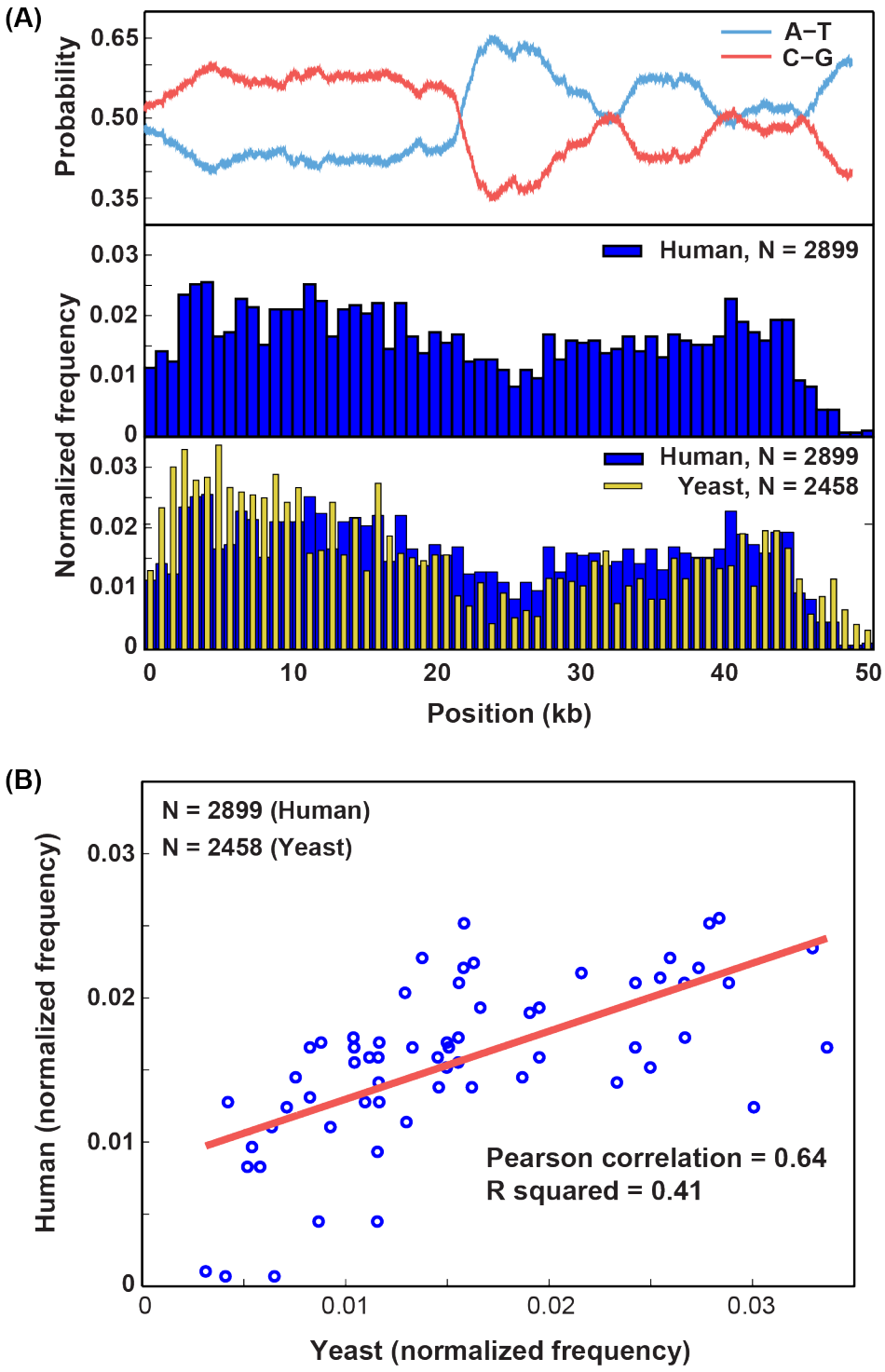
(A)
Probability
0.65
0.50
0.35
A–T
C–G
Human, N = 2899
Normalized frequency
0.03
0.02
0.01
0
Human, N = 2899
Yeast, N = 2458
0
10
20
30
40
50
Position (kb)
(B)
N = 2899 (Human)
N = 2458 (Yeast)
Human (normalized frequency)
Pearson correlation = 0.64
R squared = 0.41
Yeast (normalized frequency)

Figure 2.8:

***Figure 2.8*: Human nucleosome distribution on λ-DNA**. (A) Probability of A-T or G-C content on λ-DNA is shown in top panel. Distribution histogram of positions of human nucleosomes (middle panel, blue) displays that nucleosomes have less preference for A-T rich regions. Yeast data of the positioning distribution histogram(162) was overlaid with the human data deposited on λ-DNA (bottom). (B) The normalized frequency data of human and yeast were assessed using Pearson correlation analysis. Blue is scattered data points and red line shows the linear relationship between the two data.

Chapter 3. Intrinsically Disordered Regions Promote Resetting Steps in Eukaryotic Mismatch Repair

Credit section for this chapter 3: Yoori Kim, Christopher Furman, and Carol Manhart collaborated for this work. C.F. purified proteins, performed DNA binding, ATPase, and genetic assays. C.M. performed gel-based nuclease assays. Y.K. performed and analyzed all single-molecule assays.

Introduction

Intrinsically disordered regions (IDRs) are structurally heterogeneous protein domains that encode diverse functions. IDRs are conformationally flexible, facilitating interactions with multiple partners through intramolecular and intermolecular mechanisms (2, 167). IDRs are often found as linkers connecting functional domains where they can regulate protein stability (167). IDRs are prevalent in chromatin-binding proteins, and the IDRs in these proteins have been implicated in bridging DNA strands, chromatin remodeling, and interacting with other key proteins in DNA metabolic pathways (28, 29). Moreover, IDRs in transcription factors and single-strand DNA binding (SSB) proteins have been reported to tune the DNA binding affinities of these proteins (34, 110, 168–171). Whether these IDRs also regulate scanning on chromatin and other catalytic processes is an open question. This is partly because mutations in such regions often do not confer a specific phenotype, and in some cases, the amino acid sequences contained within IDRs, which are typically poorly conserved among family members, can be critical for the function of a specific IDR-containing protein. Using the mismatch repair protein Mlh1-Pms1 as a case study, we

explore the role of IDRs in regulating the DNA scanning and enzymatic activities of a critical eukaryotic DNA repair factor.

The MutL homolog (MLH)-family protein Mlh1-Pms1 is essential for eukaryotic DNA mismatch repair (MMR). Mlh1-Pms1 organizes into a ring-like structure that links the ordered N- and C-terminal domains via 160-290 amino acid-long IDRs (108, 109, 172–175) (Figure 3.1A). Mlh1-Pms1 searches for MutS homologs (MSH) bound to DNA mismatches (108, 163, 176). A latent MLH endonuclease activity then nicks the newly-synthesized DNA strand resulting in excision of the mismatch(177). This activity requires PCNA, and multiple nicks are thought to be necessary for the excision step of MMR (79, 80, 85, 103, 178–181).

All MLH-family proteins encode an IDR between the structured N- and C-termini. However, the functional role(s) of the IDRs in Mlh1-Pms1 is enigmatic. The composition and length of the MLH IDRs are critical for efficient MMR in yeast, and missense/deletion mutations within these linkers are found in human cancers(60, 109, 182). We previously proposed that the Mlh1-Pms1 IDRs are sufficiently long to accommodate a nucleosome within the complex, possibly allowing Mlh1-Pms1 to navigate on chromatin *in vivo*(81, 163). In support of this model, Mlh1-Pms1 foci show a dynamic behavior consistent with a rapid identification of lesions(78). In addition, the Mlh1-Pms1 IDRs display nucleotide-dependent conformational transitions, with ATP binding bringing the N- and C-terminal subunits close together(94, 97, 183). This ATPase activity is required for MMR *in vivo* and can stimulate the endonuclease activity *in vitro* (77, 92, 105, 178, 184). ATP-dependent conformational rearrangements involving the IDRs are hypothesized to position bound

DNA near the endonuclease active site and presumably change Mlh1-Pms1 affinity for DNA(94, 97, 105). Together, these studies suggest that MLH proteins may use conformational changes mediated by the ATP cycle to modulate affinity for DNA, navigate on chromatin, and introduce nicks on a DNA substrate for efficient MMR. However, these possible functions of the IDRs have not been tested directly.

Here, we use a combination of genetics, ensemble biochemistry, and single-molecule biophysics to investigate how the Mlh1-Pms1 IDRs promote both DNA scanning and nuclease activities. We show that both the sequence composition and the precise length of the IDRs are required for optimal MMR *in vivo*. Having mapped genetic requirements for MMR, we next biochemically characterized a double linker deletion (DLD) mutant that was almost completely defective in MMR (DLD_{null}), and another linker deletion mutant that retains partial *in vivo* MMR function (DLD_{MMR}). Interestingly, both mutants can bind and diffuse on DNA, retain single-round endonuclease activities, but show reduced DNA-dependent ATPase and nucleosome bypass activities. Furthermore, DLD_{null} is unable to navigate dense nucleosome arrays and is defective in multiple rounds of DNA nicking. These results establish that the IDRs license Mlh1-Pms1 to navigate chromatin and nick DNA at multiple sites to promote efficient MMR *in vivo*. They suggest that the IDRs play a critical role in regulating how a DNA repair enzyme scans chromatin for a specific target and how the enzyme activates its endonuclease activity. More broadly, these results expand the functions of IDRs in regulating the DNA scanning and enzymatic activities of chromatin-associated complexes.

Materials and Methods

Bulk biochemical assays

DNA substrates for bulk biochemical assays: pUC18 (Invitrogen) was used as the closed circular substrate for endonuclease assays presented in Figure 3.6A. A 49-mer homoduplex DNA substrate was labeled on the 5' end with ^{32}P labeled phosphate using T4 polynucleotide kinase (New England Biolabs). Unincorporated nucleotide was removed using a P30 spin column (BioRad). The two oligonucleotides were annealed by combining end-labeled AO3142 with a 2-fold molar excess of unlabeled AO3144-5'-CACGCTACCGAATTCTGACTTGCTAGGACATCTTTGCCCACGTTGACCC-3' in buffer containing 10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, and 0.1 mM EDTA. Annealing was accomplished by incubating the DNA substrates at 95 °C for 5 min, followed by cooling to 25 °C at a rate of 1 °C/min. Following annealing, excess single-stranded DNA was removed using an S300 spin column (GE). 2.7 kb pUC18 for endonuclease assays on circular DNA was purchased from Thermo. For Figure 3.6B, 4.4 kb pBR322 plasmid (Thermo) was linearized using *Hind*III (NEB) by incubation at 37 ºC for 60 min, followed by enzyme inactivation at 80 ºC for 20 min. Linearized fragments were isolated using a PCR clean-up kit (Zymo Research).

Protein purification: Yeast WT, DLD_{MMR} and DLD_{null} Mlh1-Pms1 variants were purified from galactose-induced *S. cerevisiae* BJ2168 (*MATa, ura3–52, leu2–3, 112, trp1–289, prb1–1122, prc1–407, pep4–3*) containing expression vectors as previously described(109, 185). Mlh1 contains a FLAG tag at position 499 in wild-type at the equivalent position in

Mlh1 truncation mutants. Yeast RFC and PCNA were expressed and purified from *E. coli* as described previously(186, 187). RPA-RFP was expressed and purified from Rosetta(DE3)/pLysS cells as described previously (188).

Endonuclease assay: Endonuclease reactions were performed in a buffer containing: 20 mM HEPES- KOH (pH 7.5), 20 mM KCl, 2.5 mM $MnSO_4$, 0.2 mg/mL BSA, and 1 % glycerol(189). Reactions were stopped by the addition of 0.1 % SDS, 14 mM EDTA, and 0.1 mg/mL Proteinase K (NEB). For reactions on a circular DNA substrate, products were resolved by 1.2 % agarose gel containing 0.1 μg mL^{-1} ethidium bromide, which causes covalently closed circular DNA isoforms to separate from nicked DNA product. Gels were run in 1x TAE (Tris-acetate-EDTA) at 100 V for 45 min. Negative control lanes were used as background and were subtracted out of reported quantifications. Endonuclease assays on linear substrates were carried out and stopped as described for circular DNA substrates. Denaturing agarose gels consist of 1 % (w/v) agarose, 30 mM NaCl, 2 mM EDTA pH 7.5 run in a buffer containing 30 mM NaOH and 2 mM EDTA(190). Immediately prior to sample loading, reactions were supplemented with 30 mM NaOH, 1 mM EDTA, 3 % glycerol, and 0.02 % bromophenol blue (final concentrations), heated for 5 min at 70 ºC, then cooled for 3 min on ice. Gels were run at 50 V for ~3 h. After running, alkaline agarose gels were neutralized in 0.5 M Tris base (pH 7.5) for 30 min and stained with 0.5 μg mL^{-1} ethidium bromide for ~2 h. GelEval (FrogDance Software, v1.37) was used to quantify gels.

Filter binding assay: DNA binding assays were performed as described previously (191). Briefly, 20 μL reactions containing 4 nM ^{32}P-labeled homoduplex substrate and 11 nM unlabeled homoduplex substrate were combined with increasing amounts of protein in a reaction buffer containing 20 mM Tris-HCl (pH 7.5), 0.01 mM EDTA, 2 mM $MgCl_2$, 40 μg mL^{-1} BSA, and 0.1 mM DTT. Assays with nucleotide contain 1 mM ATP. Reactions were incubated for 10 min at 30 ºC after addition of WT, DLD_{MMR} and DLD_{null} Mlh1-Pms1. Reactions were then filtered through KOH-treated nitrocellulose filters using a Hoefer FH225V filtration device for approximately 1 min. Filters were analyzed by scintillation counting to determine DNA binding efficiency.

ATPase assay: ATPase activity was determined using the Norit A absorption method as described previously(189). Briefly, 30 μL reactions contained 0.4 μM of Mlh1-Pms1 (WT, DLD_{MMR} and DLD_{null}), 100 μM [γ-32P]-ATP, 20 mM Tris, pH 7.5, 2.0 mM $MgCl_2$, 0.1 mM DTT, 1 mM $MnSO_4$, 75 mM NaCl, 1 % glycerol, 40 μg/ml BSA. Reactions were incubated for 40 min at 37 °C. When specified, DNA (49-mer homoduplex DNA substrate as described above) and PCNA were included at 0.75 μM and 0.5 μM, respectively.

Strains and plasmids: Yeast strains were grown in yeast extract/ peptone/dextrose, minimal complete, or minimal selective media (192). Full details of plasmid and strain constructions are available upon request. Expression vectors were derived from pMH1 (*GAL1-MLH1-VMA-CBD,2μ, TRP1*) and pMH8 (*GAL10-PMS1,2μ, LEU2*)(185).

Linker arm replacement series: A series of *ARS-CEN* vectors were created to test if the 50 amino acid deletion made in the Pms1 linker arm (*pms1Δ584–634*) could be replaced by

other sequences. These vectors were derived from pEAA238, which expresses *PMS1* from its native promoter (193). Vectors used to overexpress and purify Mlh1-Pms1 were derived from pMH1 (*GAL1-MLH1-VMA-CBD,2μ, TRP1*) and pMH8 (*GAL10-PMS1,2μ, LEU2*)(185). Insertion plasmids were constructed using NEB HiFi DNA Assembly cloning (pEAA644-656) and Q5 mutagenesis (pEAA659-665). The desired DNA sequence (PCR amplified from specific plasmid or constructed as gBlocks, IDT) was inserted into the deleted region (amino acids 584 to 634) of the Pms1 linker. The DNA sequence of vectors constructed using PCR amplified vector backbones and linker inserts were confirmed by DNA sequencing (Cornell BioResource Center).

lys2::insE-A_{14} reversion assay: Assays were performed by Chris Furman as described previously(109). Briefly, pEAA238 (*PMS1*), pEAA548 (*pms1Δ584-634*) and derivative linker insertion plasmids of pEAA548 were transformed into EAY3097 (*MATa, ura3–52, leu2Δ1, trp1Δ63, his3Δ200, lys2::insE-A_{14}, pms1Δ::KanMX4*) using standard methods(192, 194). Plasmids were maintained by growing strains in minimal selective histidine dropout media. When tested in combination, pEAA238, pEAA548 (*pms1Δ584-634*) or derivative linker insertion plasmids were co-transformed with pEAA213 (*MLH1*) or pEAA526 (*mlh1Δ348–373* (FLAG499)) into EAY1365 (*MATa, ura3–52, leu2Δ1, trp1Δ63, his3Δ200, lys2::insE-A_{14}, mlh1Δ::KanMX4, pms1Δ::KanMX4*). Plasmids were maintained by growing strains in minimal selective histidine and leucine dropout media. Null controls were transformed with pRS413 and pRS415 dummy vectors (195). Rates of *lys2::insE-A_{14}* reversion were calculated as $\mu = f/\ln(N \cdot \mu)$, where f is reversion frequency and N is the total number of revertants in the culture(196). For each strain, 15–45 independent

cultures, obtained from two to three independent transformants bearing a unique allele, were assayed to determine the mutation rate; 95% confidence intervals and all computer-aided rate calculations were performed as previously described(109).

Single-molecule experiments and analysis

Imaging Mlh1-Pms1 on DNA curtains: The Mlh1-Pms1 complexes used in this study contain a FLAG epitope tag at residue 499 on the Mlh1 subunit. 25 nM of FLAG-tagged proteins were conjugated with 30 nM biotinylated anti-FLAG antibody (Sigma-Aldrich, F9291-2MG) and 25 nM streptavidin QDs (Life Tech, Q10163MP) in a total volume of 60 μL on ice for 7 minutes. The mixture was supplemented with 100 μL biotin and diluted to a total volume of 150 μL in BSA buffer (40 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 0.2 mg mL^{-1} BSA, 50 mM NaCl, 1 mM DTT). The fluorescently labeled proteins were injected into the flowcell immediately after the conjugation at a 200 μL min^{-1} flow rate.

The DNA binding affinity of Mlh1-Pms1 is sensitive to salt concentration. We, therefore, used a two-step procedure to efficiently load the fluorescently-labeled protein onto DNA curtains. Mlh1-Pms1 was initially injected into the flowcell containing double-tethered DNA curtains with BSA buffer that includes low salt to assist its DNA binding. Next, the buffer was switched to imaging buffer (40 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 0.2 mg mL^{-1} BSA, 150 mM NaCl, 1 mM DTT, 1 mM nucleotides as indicated). After the flowcell was completely washed with the imaging buffer, flow was terminated to observe 1D diffusion on doubly-tethered DNA substrates.

Fluorescent labeling of nucleosomes: Nucleosomes were fluorescently labeled *in situ* after Mlh1-Pms1 diffusion trajectories were recorded on the DNA substrates. An anti-HA antibody targeting (Immunology Consultants Laboratory, RHGT-45A-Z) was diluted 100-fold in BSA buffer and injected into the flowcell at 10 nM final concentration for 5 minutes. Next, 10 nM secondary antibody was injected and incubated for 7 minutes, then buffer flow was stopped to visualize nucleosomes on double-tethered DNA molecules. We have used anti-rabbit Alexa488 (Life Tech, A-11008) or anti-rabbit ATTO647N (Sigma-Aldrich, 40839-1mL) for the secondary antibody.

Particle tracking: Fluorescently-labeled proteins were tracked in ImageJ with a custom-written particle tracking script (available upon request) and the resulting trajectories further analyzed in MATLAB (R2015a, Mathworks). The positions of labeled proteins were determined by fitting every single fluorescent particle to a two-dimensional Gaussian distribution, and the series of time-dependent sub-pixel positions generated each trajectory. Mlh1-Pms1 diffusion coefficients were determined by using the trajectories of individual moving molecules on double-tethered DNA curtains in the absence of buffer flow. The one-dimensional (1D) mean squared displacement (MSD) of each particle was determined as a function of the time interval, Δt using following equation:

$$MSD(n\Delta t) = \frac{1}{N-n}\sum_{i=1}^{N-n}(y_{i+n} - y_i)^2$$

where N is the total number of frames in the trajectory, n is the number of frames for a given time interval and ranges from 1 to N, Δt is the frame rate, and y_i is the Mlh1-Pms1 position at frame i. The MSD was calculated using the first ten time intervals (e.g. Δt =

0.05 s to 0.5 s when the frame rate was 0.05 s) and plotted as a function of Δt. Plots were fit to a line and the slope was used to calculate diffusion coefficients of individual Mlh1-Pms1 molecules. Diffusion coefficients were calculated for ≥ 30 molecules in all experiments and are reported as a mean ± standard error of the mean (S.E.M).

Measuring single nucleosome bypass frequencies: Fluorescently-labeled Mlh1-Pms1 was loaded onto double-tethered nucleosomal DNA curtains as described above. All nucleosome bypass experiments were done in imaging buffer containing 150 mM NaCl and either no nucleotide or with 1 mM ATP. We determined each collision and bypass event from individual Mlh1-Pms1 trajectories. First, a 'collision zone' was defined around each nucleosome position. Next, the positions of diffusing Mlh1-Pms1 were plotted relative to the center of the nucleosome collision zone. The number of collisions was determined by counting the number of times that Mlh1-Pms1 entered the nucleosome collision zone. Bypass events were defined as collisions that had Mlh1-Pms1 cross from the first to the second side of the nucleosome collision zone. Non-bypass events had Mlh1-Pms1 start and end the collision on the same side relative to the nucleosome. To compare the probability of bypassing single roadblock between different conditions with a statistical test, we coded each bypass event as '1' and no bypass as '0', and fit the data to a binary distribution using MATLAB.

Statistical methods: We conducted the two-sample Kolmogorov-Smirnov (K-S) test to determine whether average diffusion coefficient differ based on nucleotide types using the PAST3 software package (197). Error bars on the quantified single nucleosome bypass and

percentage of moving molecules were calculated in MATLAB using bootstrap analysis with replacement (198). P-values between conditions on single nucleosome bypass experiments were determined in MATLAB using a binary regression model. The significance threshold was set at 0.05 in all tests.

Single-molecule nicking assay: 5 nM PCNA was loaded by 1.5 nM RFC on double-tethered DNA curtain in Mlh1-Pms1 endonuclease buffer (40 mM Tris-HCl pH 8.0, 0.2 mg mL^{-1} BSA, 50 mM NaCl, 2 mM $MnCl_2$, 1 mM DTT, 1 mM ATP) (199). $MgCl_2$ was used instead of $MnCl_2$ for manganese negative control. RFC was washed out by injecting endonuclease buffer with 300 mM NaCl for 2 minutes. 20 nM Mlh1-Pms1 complexes were loaded on the PCNA-containing DNA and incubated for 20 min at 30˚ C followed by washing with 1 M NaCl for 2 min. 50 nM RPA-RFP was then injected to label any gaps larger than 10 nucleotides. For a photobleaching experiment, RPA-RFP was imaged by a 532 nm laser (100 mW at the prism face) with 250 ms exposure time. To assess RPA foci, data were collected every 5 seconds with a shutter to reduce photobleaching.

Results

The IDRs of Mlh1-Pms1 are critical for mismatch repair.

We first examined whether the IDRs of Mlh1 (~160 amino acids) and Pms1 (~290 amino acids) contain functionally important amino acids (Figure 3.1A). Our previous study established that MMR was ablated in yeast cells that lacked the Mlh1 IDR residues 348-373 and Pms1 residues 548-634 (MMR-null double-linker deletion, DLD_{null}; mlh1Δ348-

373-pms1Δ584-634). This result was surprising because deleting the same residues in the individual subunits conferred very mild MMR defects (109). Here, we expand on this early study by defining whether the composition and/or the lengths of the IDRs are critical for supporting MMR.

We first tested whether restoring the IDR of *pms1Δ584-634* to its full-length rescued MMR. *PMS1* was chosen because truncating its IDR at different positions showed only minor MMR defects and thus may be more likely to restore function with a synthetic linker (109). The substitutions included random scrambling of the 50 critical amino acids (584-634) in Pms1, as well as two biophysically characterized serine-rich regions that were equal or longer than 50 amino acids (obtained from the Herpes Virus ICP4 and *Neurospora crassa* Su9 proteins) (200, 201). All substitutions were initially examined in the wild-type (WT) *MLH1* background, where they did not restore function. The MMR defects conferred by these *pms1* mutants were similar to the *pms1Δ584-634* allele, indicating that the insertions are unlikely to disrupt the stability of the Mlh1-Pms1 complex (Figure 3.1B and Figure 3.2). In the mlh1Δ348-373 background, PMS1 linker substitutions all conferred a nearly-null MMR phenotype that was reminiscent of the DLD_{null} MMR defect. We also performed a full-length linker swap between the IDRs in MLH1 and PMS1 (Figure 3.1B and Figure 3.2); these alleles, as swaps or single substitutions, were unable to confer MMR function. Lastly, fine-scale mapping of the *PMS1* 584-634 region using scrambled and single amino acid substitution analyses identified a 20 amino acid region, 594-613, that plays a critical role for the function of the linker, and a single substitution, *pms1-Y613A*, conferred a mutator phenotype, (p-value <.00001 to WT; p-value<.00001 to *pms1Δ584-*

634 compared by Mann-Whitney U test) (Figure 3.2). Together these experiments establish that the specific sequence of the IDR, but not the flexibility, length or disorder is important for efficient MMR.

IDRs regulate Mlh1-Pms1 ATPase activity in the presence of DNA and PCNA.

Mlh1-Pms1 is a DNA-stimulated ATPase and PCNA-activated endonuclease. Nucleolytic cleavage of the newly-synthesized DNA strand by Mlh1-Pms1 is proposed to be a critical strand discrimination signal during MMR (178, 202, 203). We sought to understand the role(s) of the IDRs in promoting the enzymatic activities of Mlh1-Pms1. We compared WT Mlh1-Pms1 to two additional mutant complexes: one mostly functional in MMR (DLD_{MMR}; mlh1Δ348-373-pms1Δ437-487), and a second defective (DLD_{null}; mlh1Δ348-373-pms1Δ584-634) (Figure 3.3).

All Mlh1-Pms1 variants bound similarly to a 49 bp oligonucleotide in the absence of ATP (Figure 3.3B). In the presence of ATP, Mlh1-Pms1 displayed reduced binding to DNA, but both DLD_{MMR} and DLD_{null} displayed DNA binding levels that were higher than WT. These results show that the two DLD complexes are impaired in ATP-dependent interactions with DNA. The ATPase activities of the WT complex are stimulated by DNA and PCNA (94, 180). However, neither DLD complex exhibited such stimulation (Figure 3.3C). We conclude that the IDRs facilitate interactions between Mlh1-Pms1 and DNA, and also promote ATP hydrolysis. Remarkably, both DLD_{MMR} and DLD_{null} showed similar defects in DNA binding and ATPase activities but had very different MMR phenotypes (Figure 3.3). This puzzle encouraged us to further explore the role of the IDRs in MMR.

The IDR linker promotes facilitated 1D diffusion of Mlh1-Pms1 on DNA.

DNA-binding proteins, including Mlh1-Pms1, locate their targets using facilitated 1-dimensional (1D) diffusion along the genome (81, 163, 176, 204). Based on the biochemical results presented above, we hypothesized that the IDRs of Mlh1-Pms1 are essential for efficient 1D diffusion on chromatin. I examined Mlh1-Pms1 diffusion on double-tethered DNA curtains (Figure 3.4). In this assay, a 48.5 kb-long DNA substrate is extended over a fluid lipid bilayer between two microfabricated chromium barriers (50, 205, 206). The lipid bilayer provides a biomimetic surface that passivates the flowcell surface from non-specific adsorption by DNA-binding proteins. A single FLAG epitope was inserted at amino acid 499 of Mlh1 for downstream fluorescent labeling. The FLAG epitope does not impact Mlh1-Pms1 activities *in vitro* and *in vivo* (81, 163). For fluorescent labeling, Mlh1 was conjugated with an anti-FLAG antibody harboring a fluorescent quantum dot (QD) (81, 163). Using this assay, I characterized WT Mlh1-Pms1, as well as DLD_{MMR} and DLD_{null} variants. All three Mlh1-Pms1 complexes readily bound DNA and >90% of the molecules rapidly diffused along the entire length of the DNA substrate (Figure 3.4C; WT: 97%, N=62/64; DLD_{MMR}: 97%, N=79/81; DLD_{null}: 90%, N=60/67). Analysis of the movement showed linear mean-squared displacement (MSD) plots, verifying that all three Mlh1-Pms1 complexes freely diffuse on DNA.

ATP binding to Mlh1-Pms1 results in dimerization of the N-terminal domains, compaction of the IDRs, and the formation of a ring-like sliding clamp on DNA (87, 92, 94, 97). To probe the functional significance of this conformational change, I measured the diffusion coefficients of the Mlh1-Pms1 variants as a function of ATP. Diffusion

coefficients in the ATP-bound state were significantly increased compared to the apo (no nucleotide) condition for all complexes. These results are consistent with a prior single-molecule report of ATP-dependent diffusion of *E. coli* MutL homodimer (204). However, compared to WT and DLD_{MMR}, the mean DLD_{null} diffusion coefficient is ~six-fold lower on DNA in the presence and absence of ATP (Figure 3.4D and Table 3.1). While DLD_{null} displayed the lowest diffusion coefficient of all the complexes in the absence or presence of ATP, DLD_{null} and DLD_{MMR} displayed similar diffusion coefficients in the presence of ADP or AMP-PNP. We conclude that the IDRs of Mlh1-Pms1 are critical for efficient facilitated diffusion on DNA.

IDRs promote facilitated diffusion on nucleosome-coated DNA.

Mlh1-Pms1 must efficiently traverse chromatin to locate mismatch-bound MSH complexes. To investigate how the IDRs regulate movement on chromatin, I imaged Mlh1-Pms1 on nucleosome-coated DNA substrates. Nucleosomes were assembled using salt gradient dialysis with increasing concentrations of histone octamers to DNA molecules to recapitulate both sparse and dense nucleosome arrays (132, 163). Single nucleosomes were visualized via a fluorescent antibody directed against a triple HA epitope on the N-terminus of H2A. Nucleosomes were distributed over the entire length of the DNA molecule, with a weak preference for GC-rich segments, as described previously (Figure 2.8) (162).

I first determined whether the Mlh1-Pms1 IDRs regulate diffusion past a single nucleosome. DNA substrates with one to seven nucleosomes were assembled into double-tethered DNA curtains. Mlh1-Pms1 was added to the flowcell prior to fluorescently labeling the nucleosomes. Keeping the nucleosomes unlabeled guaranteed that Mlh1-Pms1

was not blocked by the H2A-targeting antibody. After recording 10-15 minutes of Mlh1-Pms1 diffusion, a fluorescently labeled anti-HA antibody visualized the nucleosome positions. Diffusing Mlh1-Pms1 complexes encountered and occasionally bypassed individual nucleosomes (Figure 3.6). To quantitatively determine the probability of bypassing a single nucleosome, I defined a 'collision zone' for each nucleosome which encompasses three standard deviations of the spatial resolution of our single-molecule assay (0.08 μm; ~ 300 bp) (Figure 3.5). Diffusing Mlh1-Pms1 that entered this collision zone from one side of the nucleosome and emerged from the other side was counted as a bypass event. Events where Mlh1-Pms1 entered and emerged from the same side of the nucleosome collision zone were scored as non-bypass encounters. This quantification likely underestimates the frequency of microscopic Mlh1-Pms1 nucleosome bypass events that are below our spatial resolution but does not change any of the underlying conclusions comparing the different complexes.

WT Mlh1-Pms1 bypassed nucleosomes 30 ± 0.3% of the time (Table 3.2). A molecule that travels via a 1D random walk involving facilitated diffusion has a 50% probability of stepping forward or backward on DNA. This 50% probability value is the maximum theoretical bypass probability in the absence of any nucleosome obstacles. Thus, Mlh1-Pms1 is capable of efficiently bypassing a nucleosome obstacle. In contrast, both DLD_{MMR} and DLD_{null} complexes had a 2-fold reduced nucleosome bypass frequency (18 ± 0.5%; N=29 for DLD_{MMR}; 19 ± 0.2%; N=27 for DLD_{null}).

Next, I explored how ATP-induced conformational changes affect nucleosome bypass by Mlh1-Pms1 (Figure 3.6C). In the presence of ATP, all Mlh1-Pms1 variants

exhibited a reduced bypass probability, with a significantly larger, ~2 to 3-fold, decrease in nucleosome bypass probability for both DLD variants. The decrease in bypass probabilities for DLD_{MMR} and DLD_{null} mirrors their ATPase activities (Figure 3.3C). Taken together, these data suggest that ATP-dependent dimerization of the N-terminal domains accompanied by conformational compaction of the IDRs reduces dynamic movement on nucleosome-coated DNA.

We reasoned that the combination of a reduced diffusion coefficient and less efficient nucleosome bypass observed with DLD_{null} may compromise its ability to navigate on dense nucleosome arrays. To test this, I increased the histone octamer to DNA ratio during salt dialysis to deposit >10 nucleosomes per DNA substrate (Figure 3.7B). At this high density, each nucleosome is optically indistinguishable due to the diffraction limit of light. Nonetheless, by using two-color fluorescent imaging we can still track individual diffusing Mlh1-Pms1 complexes on this nucleosome-coated DNA substrate (Figure 3.7). The 1D diffusion of all Mlh1-Pms1 complexes was restricted on this high nucleosome density substrate compared to naked DNA. Notably, while 1D diffusion coefficients of WT and DLD_{MMR} decreased by 3-fold compared to naked DNA, the DLD_{null} diffusion coefficient decreased 12-fold on this chromatinized DNA substrate (Figure 3.7E). Thus, the IDRs are important for promoting rapid facilitated diffusion on naked DNA but are especially critical for navigating on chromatin.

The IDRs are required for multiple rounds of endonucleolytic cleavage.

After MSH recognition, Mlh1-Pms1 deposits multiple nicks on the mismatch-containing DNA strand for efficient MMR (80, 207). Motivated by the importance of the

IDRs in promoting diffusion on both naked and nucleosome-coated DNA, we tested how these domains regulate Mlh1-Pms1 endonuclease activity. We first assayed the ability of Mlh1-Pms1 variants to nick supercoiled DNA in a well-established mismatch-independent endonuclease reaction (105, 178, 180, 202). This assay requires the ATP-dependent clamp loader RFC to load PCNA on the closed circle DNA substrate (208, 209). The endonuclease activity of WT Mlh1-Pms1 was indistinguishable from the DLD_{null} and DLD_{MMR} variants (Figure 3.8A). However, this assay cannot distinguish between singly- and multiply-nicked DNA substrates. This assay also cannot report the ATP dependence of the Mlh1-Pms1 endonuclease activity because ATP is required for RFC-dependent PCNA loading. To resolve these limitations, we established the alkaline gel-based and single-molecule endonucleolytic assays described below.

We directly tested the role(s) of ATP in Mlh1-Pms1 endonuclease activation on linear DNA substrates analyzed by denaturing gel electrophoresis. PCNA can thread onto the ends of linear DNA, abrogating the need for RFC and ATP (Figure 3.8B) (180, 202, 210). We observed that Mlh1-Pms1 endonuclease requires PCNA and is further enhanced by ATP binding. ATP hydrolysis was not required because ATPγS could support the reaction to the same extent or better than ATP, as suggested for the E. coli and Bacillus MutL (87, 94, 105). Although the DLD_{MMR} variant hydrolyzed linear DNA to approximately the same extent as wild-type Mlh1-Pms1, DNA degradation was attenuated with the DLD_{null} variant (Figure 3.8B). This was seen in the presence of ATP, but less so in the presence of ATPγS. In the assay in Figure 3.8B, extensive nicking on each DNA

molecule accounted for the observed substrate loss, and the reduced nicking by the DLD_{null} complex suggested another *in vivo* MMR defect.

Next, I developed a single-molecule assay to probe the limited nicking that likely occurs for DLD_{null} *in vivo*. This reaction was carried out in two steps. First, PCNA was loaded by RFC on double-tethered DNA curtains in the presence of ATP, as described previously (199). After flushing out RFC, Mlh1-Pms1 was incubated in the flowcell for 20 minutes (Figure 3.9A). PCNA and Mlh1-Pms1 were washed out by 1 M NaCl followed by injecting 50 nM RPA-RFP to visualize the ssDNA gaps made by multiple rounds of Mlh1-Pms1 endonuclease activity (Figure 3.9A-C). These ssDNA gaps arise from loss of short oligos formed by multiple nicks that are deposited in close proximity by Mlh1-Pms1. Alternatively, closely-spaced nicks may allow fraying of ssDNAs that are subsequently bound and displaced by RPA(211). Note that we would not be able to detect RPA foci if the nicks on the same strand created by Mlh1-Pms1 were far apart. I quantified the number of RPA foci per DNA and the number of RPA per focus via single-molecule photobleaching. RPA preferentially binds ~30 nt of ssDNA, but individual RPA molecules can bind ssDNA as short as 10 nucleotides (212, 213). Thus, I estimate that puncta with one RPA contain approximately 10-30 nt of ssDNA, whereas puncta with three or more RPA expose > 60 nt of ssDNA. Interestingly, DLD_{null} generated 6-fold fewer RPA foci (0.07 ± 0.02 RPA/DNA; N=307) than WT Mlh1-Pms1 (0.40 ± 0.02 RPA/DNA; N=382). In contrast, DLD_{MMR} was only mildly compromised (0.28 ± 0.02 RPA foci/DNA; N=420) compared to WT complex (Figure 3.9D). I also estimated the length of the exposed ssDNA by counting the number of RPA molecules bound on DNA. The number of RPA per focus

was comparable for DLD_{null} (1.1 ± 0.56 RPA; N=20) and DLD_{MMR} (0.9 ± 0.58 RPA; N=68) but was substantially lower than WT Mlh1-Pms1 (2.6 ± 1.2 RPA; N=79) (Figure 3.9E). These data indicate that IDRs are crucial for multiple rounds of DNA nicking during strand excision.

Discussion

All MLH proteins—from the *E. coli* MutL to the human Mlh1-Pms1—contain IDRs that link the structured N- and C-terminal domains (Figure 3.1). The importance of IDRs have been recognized in both bacterial and eukaryotic MMR, but the functions of this domain have remained elusive (108, 109). Here, we show that shortening, scrambling, lengthening, or swapping the IDRs caused mild to severe MMR defects, and even a single amino acid substitution in the IDR of Pms1, Y613A, caused an MMR defect *in vivo* (Figure 3.2). We therefore used three representative Mlh1-Pms1 complexes (WT, DLD_{MMR}, DLD_{null}) to further probe the mechanistic implications of altered IDRs.

The Mlh1-Pms1 IDRs undergo conformational changes throughout the ATP hydrolysis cycle (87, 94, 97, 183). Upon ATP binding, Mlh1-Pms1 adopts a ring-like, scrunched conformation (97). (Figure 3.10). ATP hydrolysis reverts the complex back to the extended open state where it is likely to dissociate from DNA (87, 92, 94, 97). Here, we show that the ATPase activity is disrupted when the IDRs are shortened (Figure 3.3C), indicating that disrupting this conformational cycle feeds back on the ATPase activity encoded in the structured N-terminus of both subunits. These data motivated us to assay the roles of the IDRs in scanning DNA and nucleolytically processing the DNA.

Mlh1-Pms1 locates MMR foci after MSH complexes recognize a DNA lesion *in vivo* (78). Therefore, Mlh1-Pms1 must scan the genome as nucleosomes are being assembled onto the newly synthesized DNA. Strikingly, the DLD_{null} complex is significantly impaired in 1D diffusion on naked DNA and this defect is further exacerbated on dense nucleosome arrays, where the diffusion coefficient of DLD_{null} is decreased by 12-fold compared to that of WT Mlh1-Pms1 (Figure 3.7). The different activities of DLD_{null} and DLD_{MMR} suggest that the residues spanning the 584-634 aa region in Pms1 are especially critical for MMR. These residues likely contribute to the conformational rearrangement of the entire complex. A second possibility is that the IDR reorganizes how DNA is channeled through the Mlh1-Pms1 complex. Further structural and biophysical studies will be required to probe the conformational transitions of these IDR variants on DNA. Taken together, our data establish that the IDRs regulate facilitated diffusion of Mlh1-Pms1 on both naked and nucleosome-coated DNA substrates.

Recent studies suggest that iterative nicking by multiple Mlh1-Pms1 complexes promote MMR by increasing the entry sites for EXO1 or stimulating EXO1-independent repair via strand displacement and/or exonucleolytic activities of Polymerase δ(79, 80, 85, 179, 207). The IDRs may control this activity by regulating the facilitated diffusion of Mlh1-Pms1 on DNA and by ATP-dependent conformational rearrangements that bring the DNA strand close to the nuclease active site. Indeed, ATP-dependent structural rearrangement stimulates the nuclease activity in the bacterial MutL system(87, 105). Consistent with this idea, DLD_{null} was defective in carrying out multiple rounds of DNA

cleavage, as seen in both ensemble and single-molecule nuclease assays (Figure 3.8 and Figure 3.9).

Together with prior studies of nucleotide-dependent MLH conformations (79, 97, 108, 183), we propose that the IDRs undergo structural rearrangements that change how the complex interacts with the DNA substrate (Figure 3.10). Mlh1-Pms1 rapidly diffuses on nucleosome-coated DNA in search of lesion-bound MSH complexes. When bound to ATP, the Mlh1-Pms1 acts as a stable sliding clamp on DNA and its endonuclease is activated by PCNA to nick DNA. Upon ADP and P_i release Mlh1-Pms1 enters the apo state and cycles off of DNA. The ATP binding/hydrolysis cycle is repeated, with a Mlh1-Pms1 sliding clamp activated by PCNA to nick DNA at a different position. Consistent with this and variants of this model (79), DLD_{null} shows *specific* defects in dynamic processes requiring a functional ATP binding/hydrolysis cycle on long DNA substrates; e.g., diffusion on naked and nucleosome coated DNA (Figure 3.4D and Figure 3.7E) and making multiple nicks (Figure 3.8 and Figure 3.9). DLD_{MMR} and DLD_{null} complexes show similar defects in single step processes; e.g., DNA binding and ATPase activity on short oligonucleotides, single nicking (Figure 3.8A), and bypassing a single nucleosome (Figure 3.6). This model also explains why locking either DLD_{null} or DLD_{MMR} in a single configuration by adding ADP or AMP-PNP results in similar diffusion kinetics. The combination of ATPase, facilitated diffusion, and endonuclease defects lead to the MMR catastrophe observed for the DLD_{null} mutant. A subset of these phenotypes explains the partial MMR defects of the other IDR variants that we assayed genetically (Figure 3.1B and Figure 3.2).

Conformational rearrangements allow Mlh1-Pms1 to rapidly diffuse on DNA. Facilitated diffusion accelerates the search for MSH-bound lesions on chromatin. Importantly, the ATP hydrolysis cycle confers dynamic structural rearrangements that stimulate multiple nicks in the vicinity of the mismatch. More broadly, our results highlight that conformational changes in intrinsically disordered linkers can profoundly alter DNA interactions and enzymatic activities of neighboring structured domains. This work adds additional details to the emerging disorder-function paradigm emerging from biophysical studies of intrinsically disordered proteins.

Table 3.1. Nucleotide-dependent Mlh1-Pms1 diffusion coefficients

Nucleotide type	Protein type	1D diffusion ± S.E.M. ($\mu m^2 s^{-1}$)	Number of molecules	p-value (relative to WT)	
				t-test	K-S test
None	WT	0.507 ± 0.07	56	N/A	
	DLD_{MMR}	0.223 ± 0.04	41	0.1	0.01
	DLD_{null}	0.078 ± 0.01	39	$1.5x10^{-6}$	$6.7x10^{-9}$
ADP	WT	0.584 ± 0.07	46	N/A	
	DLD_{MMR}	0.303 ± 0.04	40	0.001	0.0007
	DLD_{null}	0.376 ± 0.14	34	0.15	0.001
AMP-PNP	WT	0.753 ± 0.18	30	N/A	
	DLD_{MMR}	0.451 ± 0.11	41	0.13	0.99
	DLD_{null}	0.275 ± 0.05	26	0.02	0.07
ATP	WT	0.918 ± 0.10	59	N/A	
	DLD_{MMR}	0.646 ± 0.10	50	0.069	0.007
	DLD_{null}	0.171 ± 0.02	42	$8.2x10^{-8}$	$9.3x10^{-13}$

All data points were acquired in imaging buffer containing 150 mM NaCl.

Table 3.2. Probability of single nucleosome bypass by Mlh1-Pms1 complexes

Nucleotide type	Protein-Condition	Number of trajectories	Number of collisions	Probability ± std. dev.	P-value (relative to WT)	P-value (relative to DLD_{MMR})	Average number of collisions per trajectory ± std. dev.
Minus	WT	31	1361	0.30 ± 0.003	N/A		44 ± 44
	DLD_{MMR}	29	1166	0.18 ± 0.005	$2.8x10^{-11}$		40 ± 29
	DLD_{null}	27	1033	0.19 ± 0.002	$6.1x10^{-9}$	0.54	38 ± 24
ATP	WT	34	1223	0.21 ± 0.007	N/A		35 ± 35
	DLD_{MMR}	30	1572	0.063 ± 0.004	$4.8x10^{-28}$		56 ± 60
	DLD_{null}	30	1111	0.070 ± 0.002	$1.9x10^{-19}$	0.08	37 ± 30

P-values are determined from fitting binary logistic regression relative to WT with same nucleotide condition.

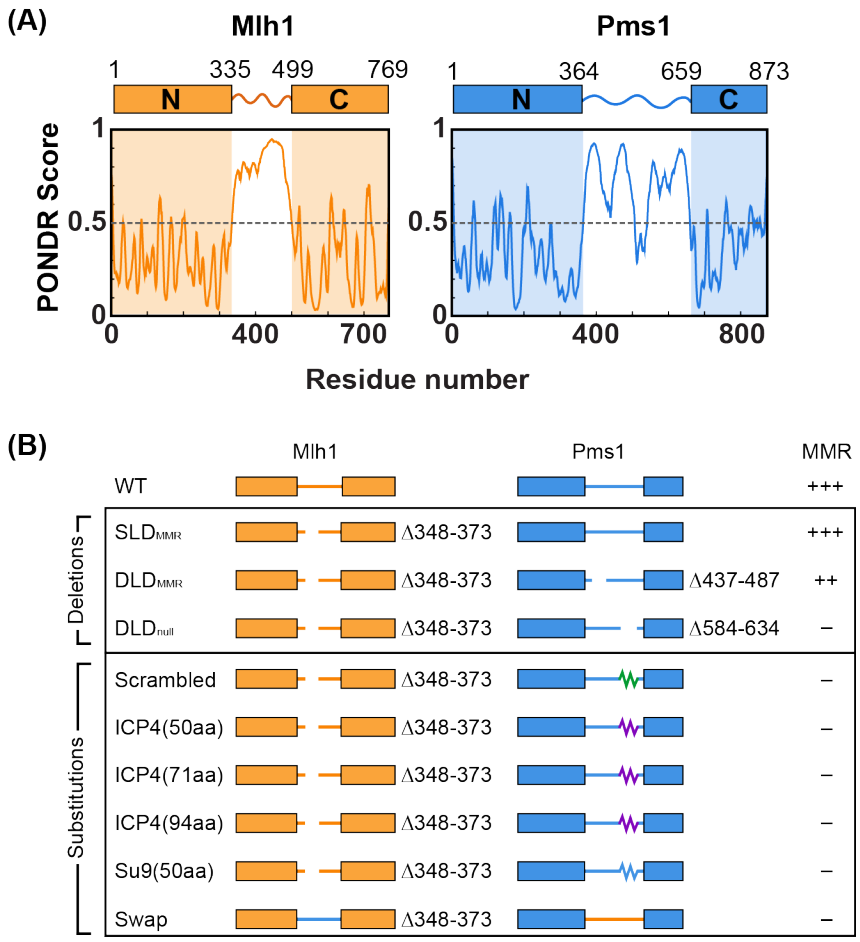

***Figure 3.1*: The IDR of Mlh1-Pms1 is critical for MMR *in vivo*.** (A) Structural prediction of long IDRs in both Mlh1 (amino acids 335-499) and Pms1 (amino acids 364-659) (214). Any value above 0.5 is considered disordered. (B) Schematic of IDR sequence changes made in Mlh1-Pms1, followed by the mutator phenotype conferred by the indicated alleles. +++ wild-type mutation rate, ++ hypomorph, – null. See text for a description of the specific sequences.

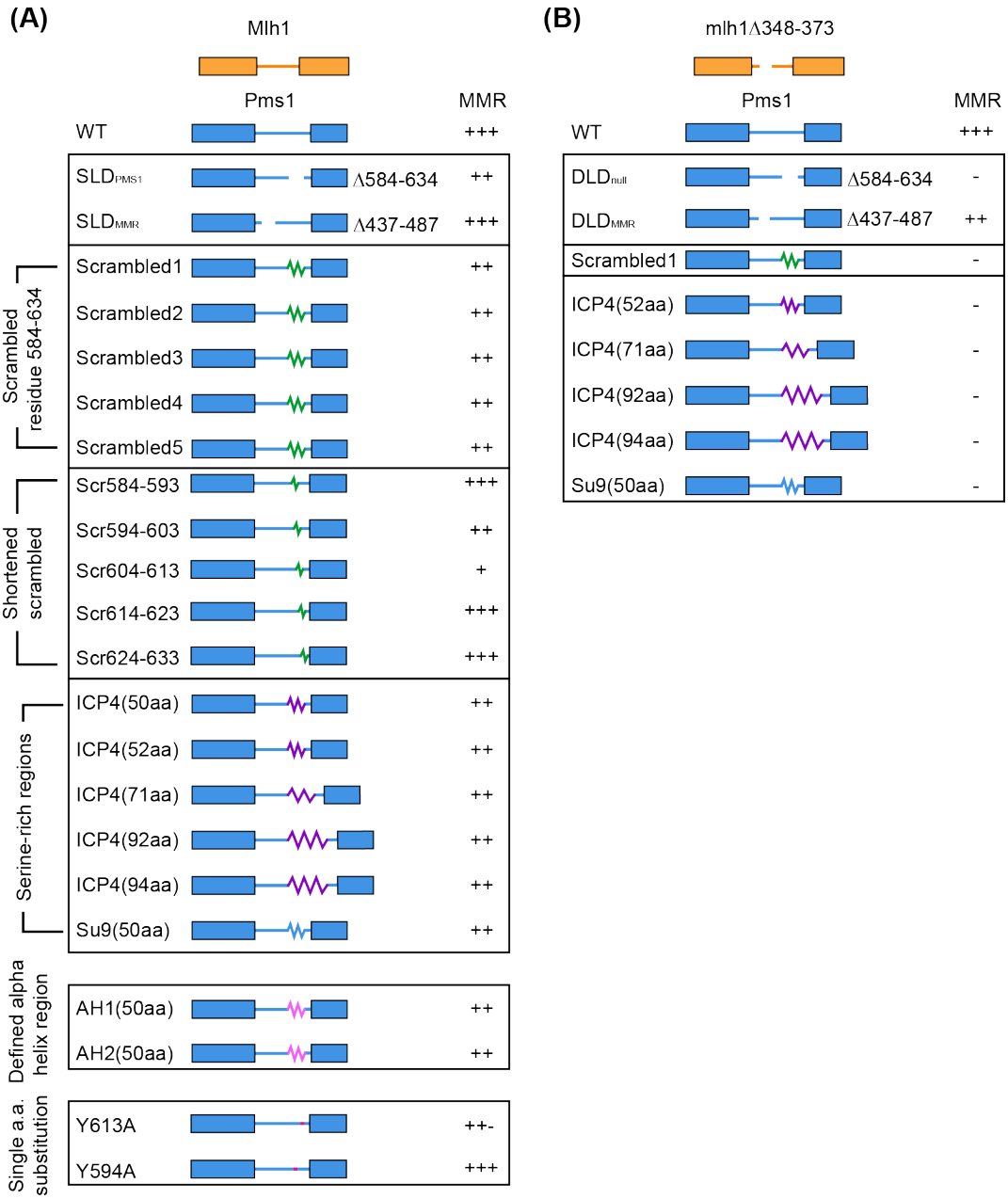

***Figure 3.2*: *In vivo* MMR assay for various linker variants of Mlh1-Pms1.** (A) Schematic of specific sequences that replace the IDRs in Pms1, followed by the mutator phenotype conferred by the indicated alleles in the *MLH1* strain background. +++ indicates a wild-type mutation rate, ++, ++- and + indicate hypomorph phenotypes, and – indicates a null phenotype. (B) Analysis of linker alleles presented in panel A in the *mlh1-Δ348-373* background.

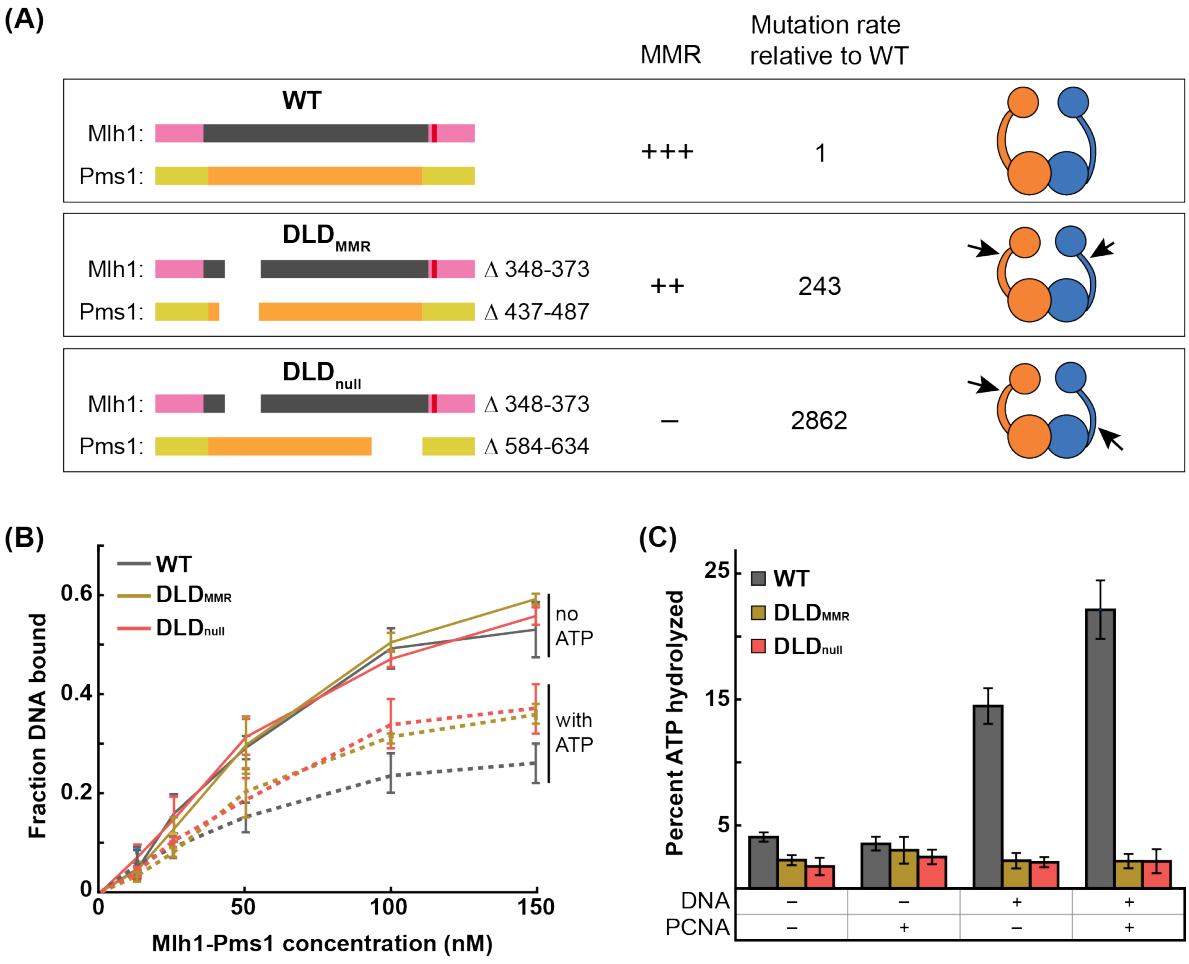

***Figure 3.3*: ATPase dependent DNA binding of Mlh1-Pms1.** (A) Schematic of two double linker deletion mutants used *in vitro* study. (B) DNA binding activities for each complex analyzed by filter binding in the presence (dashed line) and absence (solid line) of 1 mM ATP. Mlh1-Pms1 variants were included at final concentrations of 12.5 nM, 25 nM, 50 nM, 100 nM, and 150 nM in buffer containing 25 mM NaCl. DNA binding of a 49 bp oligonucleotide was quantified by scintillation counting. Three replicates were averaged; error bars indicate ± one SD. (C) ATP hydrolysis activities of WT and mutant Mlh1-Pms1 complexes (0.40 μM) were determined alone, and in the presence of PCNA (0.5 μM), or 49-bp homoduplex DNA (0.75 μM), and both PCNA (0.5 μM) and 49-bp homoduplex DNA (0.75 μM). Error bars indicate ± one SD of three replicates.

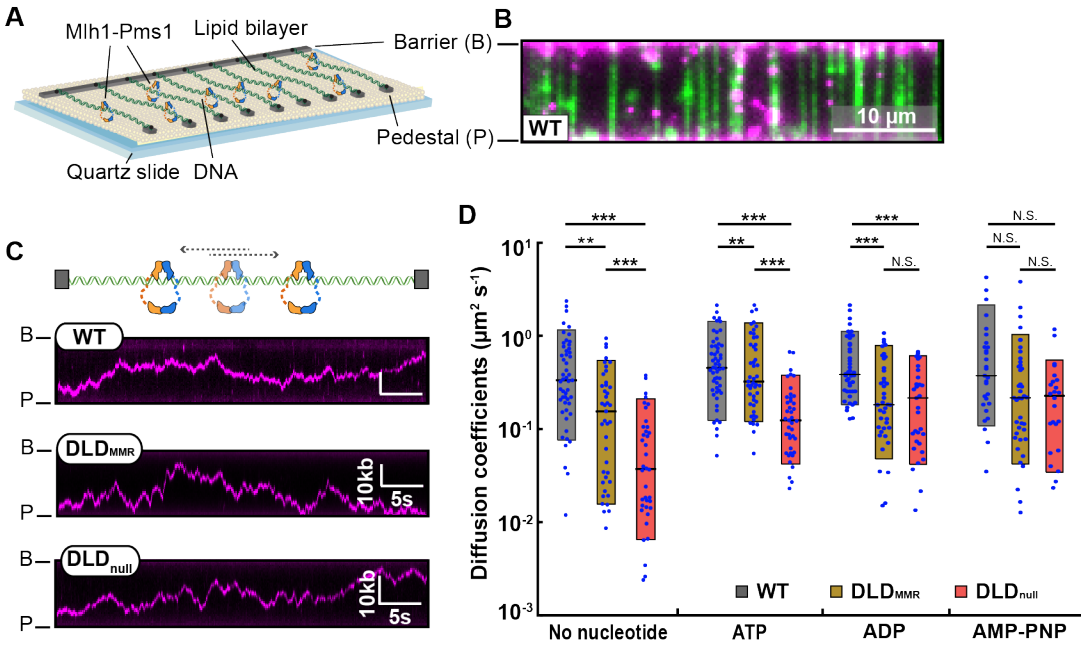

***Figure 3.4*: The IDRs promote facilitated Mlh1-Pms1 diffusion on DNA.** (A) Schematic of the DNA curtains assay. Fluorescently-labeled Mlh1-Pms1 is injected into the flowcell and visualized on double-tethered DNA substrates in the absence of buffer flow. (B) An image of Mlh1-Pms1 (magenta puncta) on double-tethered DNA molecules (green). To avoid interference from the DNA-intercalating dye, the DNA is not fluorescently stained during analysis of Mlh1-Pms1 movement on DNA. (C) A schematic (top) and representative kymographs of a WT, DLD_{MMR}, DLD_{null} Mlh1-Pms1 diffusing on DNA. (D) Diffusion coefficients of Mlh1-Pms1 complexes in different nucleotide conditions. Boxplots indicate the median, 10th, and 90th percentiles of the distribution. P-values are obtained from K-S test: * *P*-values <0.05, ** *P*-value < 0.01, and *** *P*-value < 0.005.

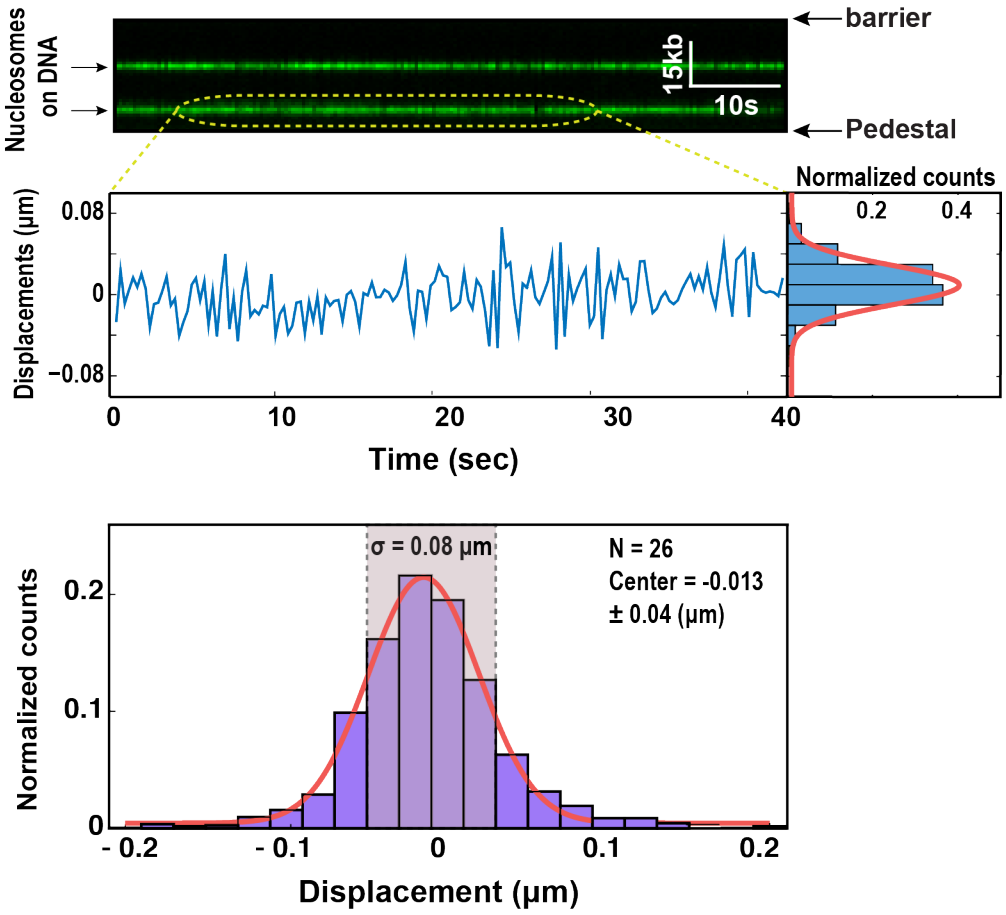

Figure 3.5: **Definition of 'nucleosome zone' for bypass analysis**. To determine the spatial resolution, a distribution of the net displacement of single nucleosomes was fit to a Gaussian distribution (red line). For analyzing single nucleosome bypass frequencies, the nucleosome zone was defined as a three-sigma region surrounding the mean nucleosome position (bottom).

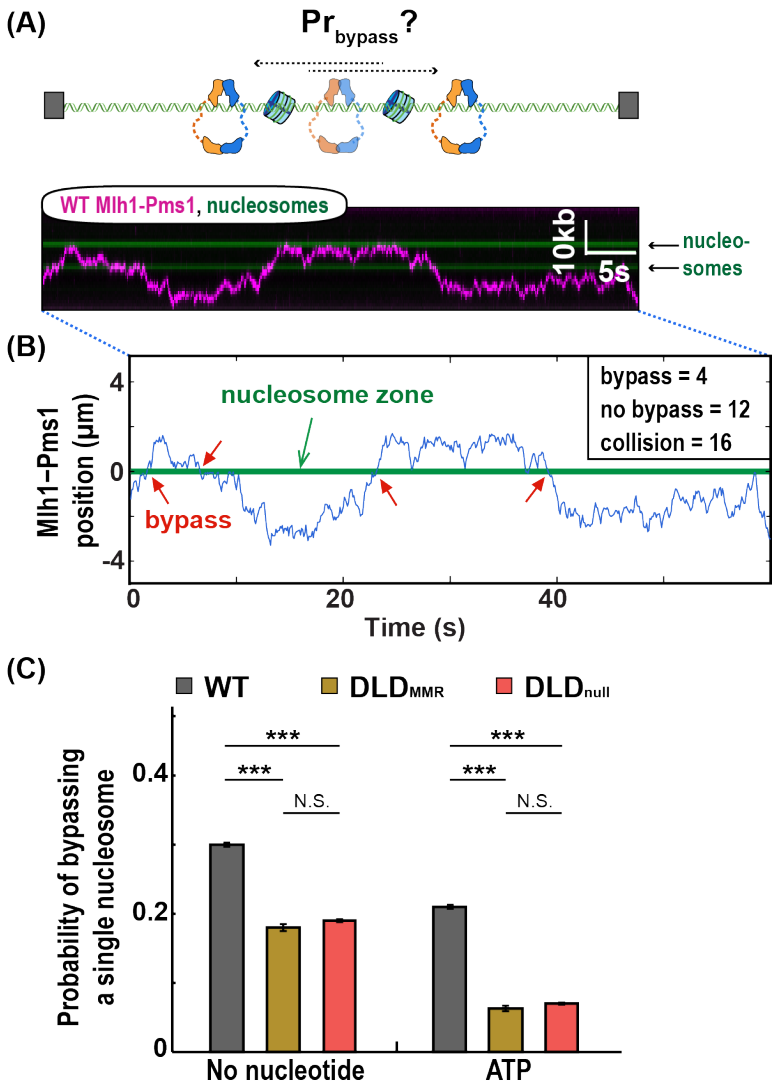

***Figure 3.6*: Analysis of nucleosome bypassing activity for three Mlh1-Pms1 complexes.** (A) An illustration (top) and a representative kymograph of WT Mlh1-Pms1 diffusing past a nucleosome (bottom). (B) Trajectory analysis of single nucleosome bypass events. (C) The values obtained from the analysis shown in (B) are fit to binary logistic regression to obtain predicted probability of nucleosome bypass.

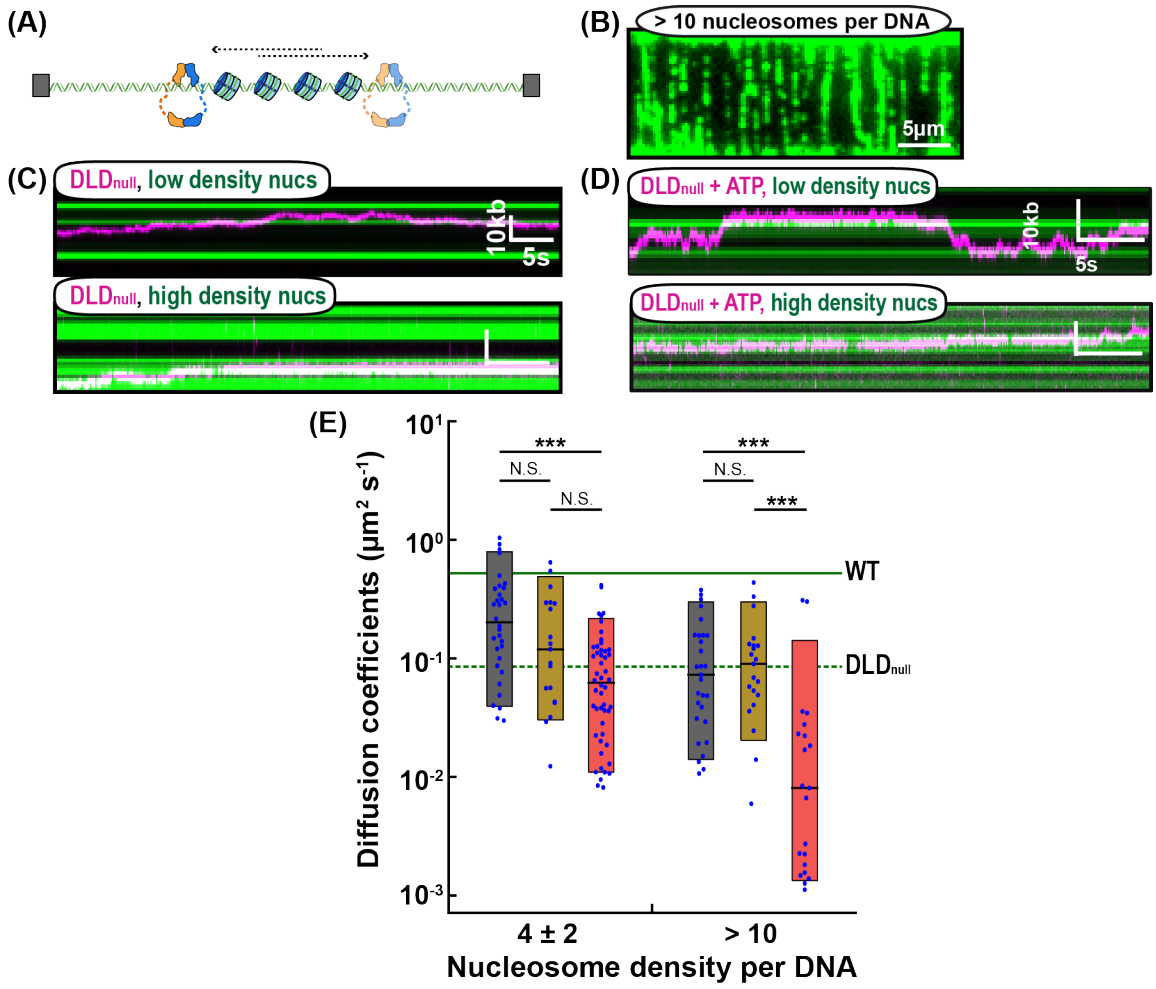

***Figure 3.7*: The IDRs increase Mlh1-Pms1 movement on nucleosome-coated DNA.** (A) A cartoon (top) of Mlh1-Pms1 movement on nucleosome-coated DNA. (B) A fluorescent image of double-tethered DNA curtain with > 10 nucleosomes (green) per DNA. The position of each nucleosome cannot be determined due to overlapping fluorescent nucleosome signals. Representative kymographs of DLD_{null} (magenta) on DNA containing 4 ± 2 nucleosomes (top, green) or > 10 nucleosomes (bottom, green) in the absence of ATP (C) or with ATP (D). (E) Diffusion coefficients of the three Mlh1-Pms1 complexes on nucleosome-coated DNA. The solid and dashed green lines indicate the mean of the diffusion coefficients of WT and DLD_{null} on naked DNA, respectively. P-values are obtained from K-S test: * *P*-values <0.05, ** *P* value < 0.01, and *** *P*-value < 0.005. N.S. indicates p > 0.05.

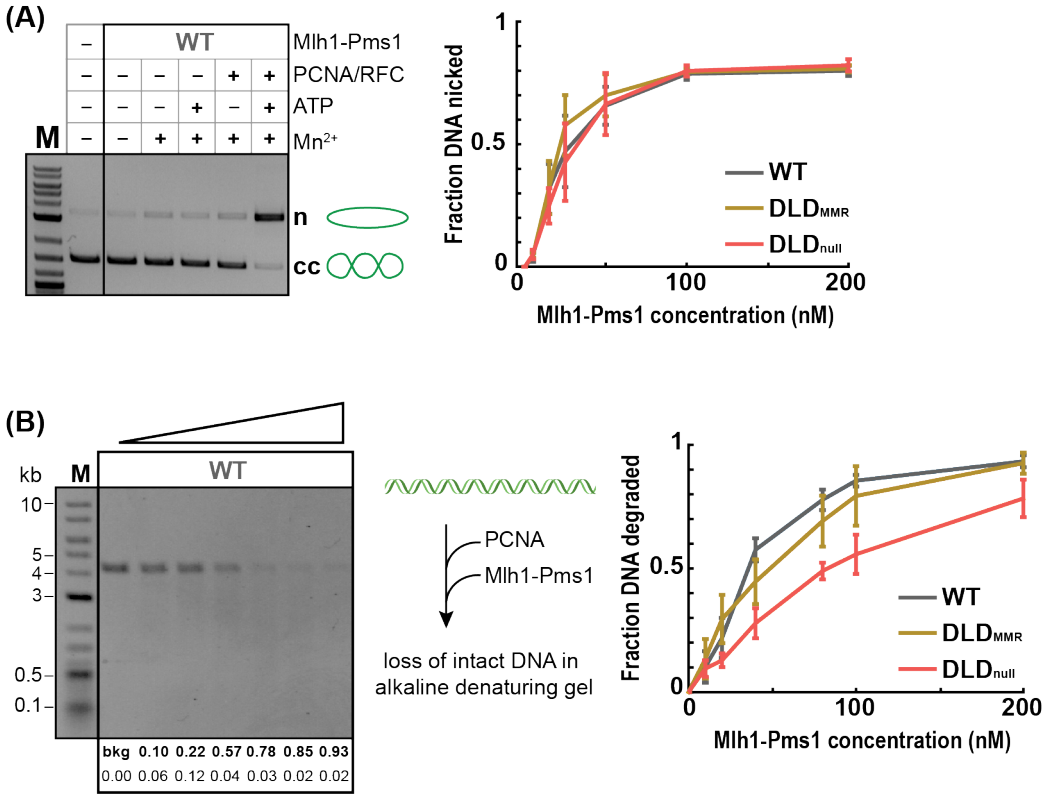

***Figure 3.8*: Bulk endonuclease assay for Mlh1-Pms1.** (A) Endonuclease activity on closed circular DNA in the presence (+) or absence (-) of $MnSO_4$, ATP, and yeast PCNA/RFC (left panel). Where + is indicated, the concentration of $MnSO_4$ was 2.5 mM, ATP was 0.5 mM, RFC and PCNA were each 500 nM. The final concentration of WT Mlh1-Pms1 was 100 nM. In the presence of $MnSO_4$, ATP, RFC, and PCNA at the above concentrations, Mlh1-Pms1 variants were titrated from 0-200 nM (right panel). Error-bars: SD of three replicates. (B) The amount of substrate lost was quantified on a single gel and expressed as a fraction by comparing to lane 2. The average fraction lost and SD were calculated from four replicates (left). Illustration (middle panel) and quantification (right panel) of the endonuclease activity of wild-type and mutant Mlh1-Pms1 complexes (titrated from 0-200 nM) on linear DNA. All reactions contain 500 nM PCNA, 0.5 mM ATP, and 5 mM $MnSO_4$. Error-bars: SD of four replicates.

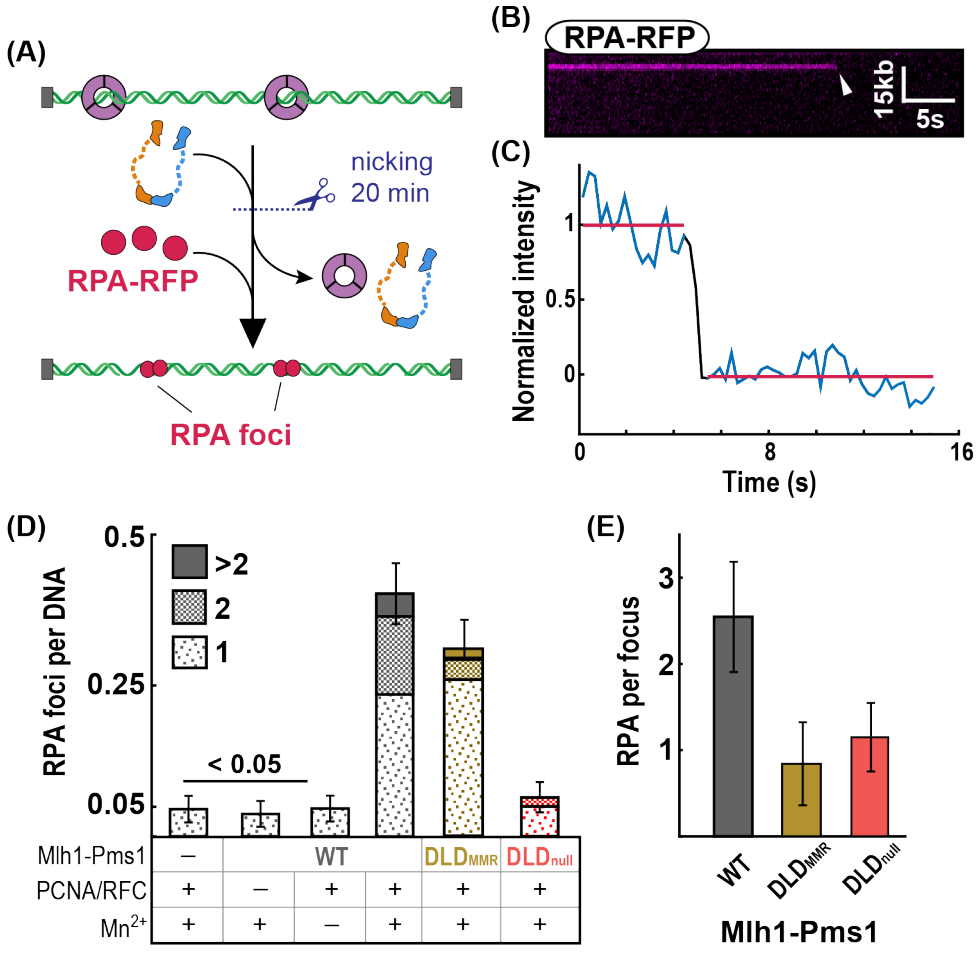

***Figure 3.9*: Single-molecule endonuclease assay for Mlh1-Pms1.** (A) Schematic of the single-molecule endonuclease assay. Formation of ssDNA gaps via PCNA-activated Mlh1-Pms1 nuclease activity was visualized by injecting RPA-RFP into the flowcell. (B) Kymograph and (C) fluorescent intensity profile of an RPA-RFP punctum with a single-step photobleaching event (arrow), indicating a single RPA-RFP molecule on the ssDNA. (D) The number of RPA foci per DNA molecule for the indicated Mlh1-Pms1 variants. (E) The number of RPA molecule per punctum for the three Mlh1-Pms1 complexes. To estimate the number of RPA molecules per ssDNA segment, the fluorescent intensity for each punctum was measured and normalized to that of a single RPA-RFP.

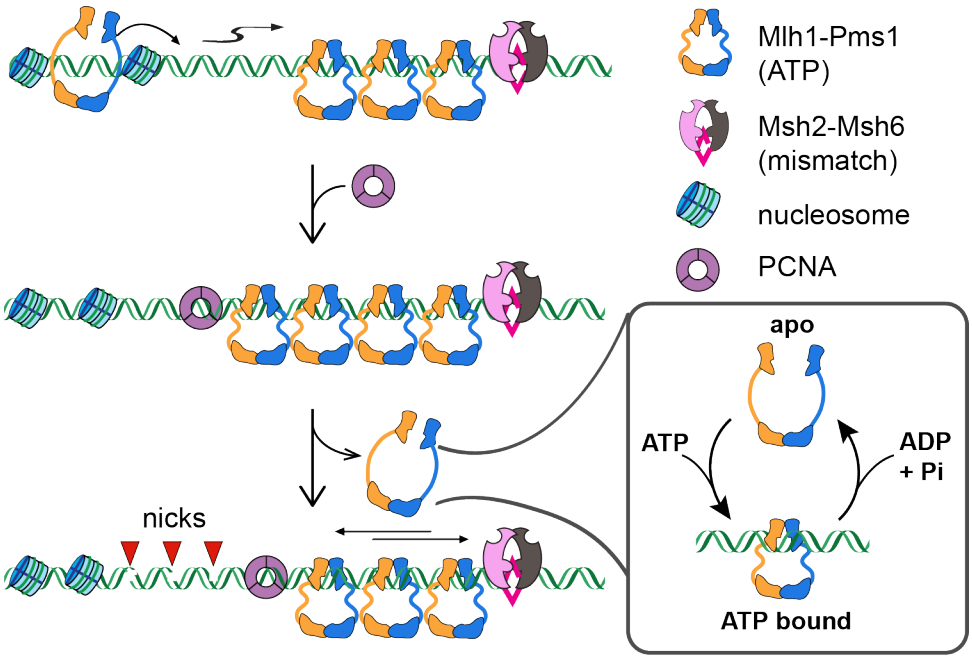

***Figure 3.10*: A model of the conformational cycle of Mlh1-Pms1 in MMR.** Mlh1-Pms1 forms open or closed sliding clamps in the absence (apo) and presence of ATP, respectively (97). The IDRs in Mlh1-Pms1 are critical for transmitting ATP binding and hydrolysis states in the N-terminal domains to the C-terminal endonuclease domain. Mlh1-Pms1 loads onto DNA where it rapidly diffuses on DNA and efficiently bypasses nucleosome barriers in search of Msh complexes bound to mismatches. In the ATP bound form, the complex forms a sliding clamp, and its endonuclease is activated by PCNA to nick DNA. After catalyzing one or more nicks, the Mlh1-Pms1 then enters the apo conformation and cycles off of DNA. Another complex in the ATP bound sliding clamp mode is then activated by PCNA to further nick the DNA, repeating the cycle.

Chapter 4. Efficient Modification of λ-DNA for Single-Molecule Study and PCNA Dynamics on Trinucleotide Repeat Structure

Portion of this chapter has been published in the following journal:

Kim,Y., de la Torre,A., Leal,A.A. and Finkelstein,I.J. (2017) Efficient modification of λ-DNA substrates for single-molecule studies. Scientific Reports, 7, 2071

– *Y.K. and A.T. contributed equally. Y.K. purified proteins, performed and analyzed all single-molecule experiments for PCNA. A.T. optimized the recombineering method. Y.K., A.T., and I.J.F. co-wrote the paper.*

Introduction

Single-molecule studies of protein-nucleic acid interactions frequently require site-specific modification of long DNA substrates. The bacteriophage λ is a convenient source of high quality long (48.5 kb) DNA. The long genomic λ-DNA is useful for observing proteins that package, transcribe, replicate, or diffuse on DNA (215). Moreover, the *cosL* and *cosR* ssDNA overhangs facilitate direct ligation of modified DNA handles. In addition, high-quality recombinant DNA can be purified from lysogenic cells in large quantities.

Site-specific λ-DNA modification strategies generally fall within one of four categories: (i) restriction enzyme cleavage and ligation (216, 217); (ii) recombinase-mediated modifications (218, 219); and (iii) insertion of extrahelical structures via oligonucleotide mimics, or (iv) nicking endonuclease (nickase)-based oligo replacement (220–224). To date, restriction enzyme cleavage and multi-step ligation *in vitro* is one of the most frequently used methods for modifying λ-DNA. However, this approach is technically challenging because multi-step intra-molecular ligation is inefficient.

Restriction enzyme-based cloning is also limited due to the few unique restriction sites within the 48.5 kb phage genome. Furthermore, producing appreciable quantities of recombinant DNA requires packaging the λ-DNA into phage particles, several rounds of viral amplification, and infection on an *E. coli* host with high-titer λ-phage prior to DNA purification.

Here, we developed a method for rapidly modifying and purifying recombinant λ-phage DNA. We used *in vivo* recombineering to target any segment of the phage genome, abrogating the need for restriction sites and ligation. Using this approach, we develop a molecular toolkit for inserting exogenous DNA sequences into the λ-phage genome with >90% efficiency. I demonstrate a strategy for inserting non-replicative extrahelical DNA structures at these sites. I explore the utility of these DNA structures by demonstrating site-specific loading of yPCNA by the clamp-loader complex Replication Factor C (RFC). Our results show that yPCNA can be loaded on both flaps and a $(CAG)_{13}$ triplet nucleotide repeat (TNR). Surprisingly, yPCNA remains trapped within the $(CAG)_{13}$ repeat, adding further support to a model that suggests yPCNA participates in TNR expansion via illegitimate activation of the DNA mismatch repair machinery (225–227). In sum, we anticipate that this molecular toolkit will be broadly useful for both ensemble and single-molecule studies that require site-specific modification of long DNA substrates.

Materials and Methods

Preparation of λ-phage lysogens for insertion of extrahelical structures

The *E.coli* lysogen used to insert flap- or $(CAG)_{13}$-containing structures was generated by Red-based in vivo recombination (199). First, lysogens containing wild-type λ-phage DNA (IF189) were transformed with pKD78, which harbors the Red recombineering system under the control of an arabinose-inducible promoter (228). The insertion DNA including a nicking cassette were PCR amplified from the helper plasmid using *Taq* DNA polymerase (NEB# M0320S). The gel-extracted PCR products was re-suspended in Milli-Q water to a final concentration of 100-150 μg μL^{-1} and used as the targeting DNA for the recombineering reaction. Fresh electrocompetent cells were prepared for every recombineering reaction. A 5 mL LB culture of strain IF189 transformed with pKD78 was grown overnight at 30°C in the presence of 10 μg μL^{-1} chloramphenicol. The following day, 350 μL of cells were used to inoculate a fresh 35 mL culture of LB containing the same concentration of antibiotic. When the cells reached an $O.D._{600}$ ~0.5, the Red recombinase system was induced by adding 2% L-arabinose (GoldBio) and incubated for an additional 1 hour at 30˚C. Cells were harvested at 4,500 RCF for 7 min, washed three times in ice-cold Milli-Q H_2O, and finally resuspended in 200 μL of H_2O (229). For recombineering, 50-150 ng of targeting DNA was electroporated at 18 kV cm^{-1} in 0.1 cm cuvettes using a micropulser (Biorad #165-210). Cells were immediately resuspended in 1 mL of SOC and then transferred to culture tubes containing 10 mL LB broth. After 4 hours outgrowth at 30°C, 100 μL of the culture was plated onto LB agar plates containing either 30 μg mL^{-1} carbenicillin or 30 μg mL^{-1} kanamycin. Colonies were checked for successful incorporation of recombinant DNA via colony PCR.

Purifying Phage DNA from lysogens

A single lysogenic colony was grown in 50 mL of LB broth with the appropriate antibiotic overnight at 30°C. 5 mL of this starter culture was used to inoculate 500 mL of LB the following morning. When the $O.D._{600}$ reaches ~0.6, the temperature was rapidly raised to 42°C in a water bath. The culture was placed at 45°C in a shaking incubator for 15 minutes and then transferred to a 37°C for two hours. To liberate the phage particles, cells were harvested by centrifugation at 3,000 RCF for 30 minutes and lysed via re-suspension in 10 mL of SM buffer (50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 8 mM $MgSO_4$) + 2% chloroform, and rotated at 37°C for 30 min. A subsequent 1 hour incubation with 50 ng μL^{-1} DNaseI (Sigma# D2821) and 30 ng μL^{-1} RNaseA (Sigma# R6513) degraded the bacterial genomic DNA and RNA. The clarified lysate, containing soluble phage capsids, was obtained by centrifugation for 15 minutes at 6,000 RCF and 4°C, and further diluted with 40 mL of SM buffer. Phage capsids were precipitated by incubating for 1 hour with 10 mL ice-cold buffer L2 (30% PEG 6000, 3M NaCl) and then harvest by centrifugation at 10,000 RCF for 10 minutes at 4°C. The phage pellet was washed with 1 mL of buffer L3 (100 mM Tris-HCl [pH 7.5], 100 mM NaCl, 25 mM EDTA) and then re-suspended with 3 mL of buffer L3, followed by an equal volume of buffer L4 (4% SDS). The phage capsid proteins were further digested by incubation with 100 ng μL^{-1} of proteinase K for 1 hour at 55°C. 3 mL buffer L5 (3 M potassium acetate [pH 7.5]) was added, and the cloudy solution was clarified by centrifugation at 15,000 RCF for 30 minutes at 4°C. The soluble phage DNA was passed over a pre-equilibrated Qiagen tip-500 column (Qiagen #10262), washed with buffer QC (1.0 M NaCl, 50mM MOPS [pH 7.0], 15% isopropanol) and eluted with 15 mL buffer QF (1.25 M NaCl, 50 mM Tris [pH

8.5], 15% isopropanol). Finally, the DNA was precipitated with the addition of 10.5 mL of 100% isopropanol, rinsed in 70% ethanol and re-dissolved in TE buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA) to a final DNA concentration of 200-500 ng μL^{-1}. We routinely obtained ~250 μg of pure λ-DNA from a single purification.

Inserting synthetic oligonucleotides into λ-DNA

25 μg of the DNA was incubated with 150 U of Nt.BspQI (NEB) in a 250 μL reaction with 1X buffer 3.1 at 55°C for 1 hour. The reaction was halted with 1 U of proteinase K (NEB #P8107S) for 1 hour at 55°C. The nicked DNA was mixed with a 100-fold molar excess of the desired insert oligo (AD006 for 5'-flap, YK105 for $(CAG)_{13}$, MB32 for mock insert, Table 4.1), along with a 10-fold excess of cosL and cosR-complementary oligos (IF003 and IF004 for cosL → cosR, IF001 and IF002 for cosR → cosL, Table 4.1). The solution was heated to 70˚C for 15 minutes followed by slow cooling to 15°C in a thermocycler at a rate of -0.5°C min-1. The annealed mixture was supplemented with 6000 U of T4 DNA ligase (NEB#M0202L) and 1 mM ATP and further incubated overnight at room temperature in a final reaction volume of 300 μL. A 50 μL aliquot was taken for alkaline agarose gel and restriction enzyme digest analysis. The remaining 250 μL was supplemented with high salt (1M NaCl) and then passed through a 120 mL Sephacryl S-1000 column (GE # 17-0476-01), in TE running buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA) plus 150mM NaCl, to separate the modified λ DNA from excess oligos and enzymes (Figure 4.1C).

Purification of *S. cerevisiae* Proliferating Cell Nuclear Antigen (yPCNA)

A plasmid expressing *Saccharomyces cerevisiae* His_6-PCNA was kindly provided by Francisco Blanco(230). A triple FLAG epitope was introduced at the N-terminus via inverse PCR mutagenesis (NEB) using primers YK_PCNA01 and YK_PCNA02 (see Table 4.1) to generate pIF105. pIF105 was transformed into BL21(DE3) codon plus RIL cells. A colony was inoculated into 30 mL LB with 50 μg mL^{-1} kanamycin and 34 μg mL^{-1} chloramphenicol, and grown overnight at 37°C. Ten mL of the overnight culture were seeded into 1 L LB and grown in the presence of both antibiotics. When the culture reached an OD_{600}~0.5, 0.8 mM IPTG was added and induction continued at 37°C for 4 hours. Cells were harvested by centrifugation at 3,000 RCF for 10 minutes, and resuspended in 50mL lysis buffer (50 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.5 mM TECP, 10% (v/v) glycerol) with 1x HALT protease inhibitor. Cells were lysed by sonication on ice and centrifuged at 95,000 RCF for 30 minutes. Imidazole was added to the supernatant to a final concentration of 30 mM. A 5 mL HisTrap HP column (GE Healthcare) was pre-equilibrated with 50 mL lysis buffer and the lysate was loaded onto the column and washed with 50 mL of Ni-buffer (50 mM Tris-HCl [pH 7.6], 150 mM NaCl, 10% glycerol (v/v), 30 mM imidazole). PCNA was eluted with a ~110 mL gradient to Ni-buffer + 500 mM imidazole over 120 minutes. PCNA-containing fractions were identified via 12% SDS-PAGE, dialyzed into storage buffer (20 mM Tris-HCl [pH 7.6], 100 mM NaCl, 10% (v/v) glycerol, 1 mM DTT) and concentrated using a centrifugal filter (10 kDa Amicon, Millipore). Small aliquots were frozen in liquid nitrogen and stored at -80°C.

Purification of *S. cerevisiae* Replication Factor C (yRFC)

The plasmids pLant2b-RFC-AE and pET11-RFC-BCD were kindly provided by Manju Hingorani(231). The two plasmids were co-transformed into BL21(DE3) ArcticExpress cells (Agilent). A single colony was inoculated into 680 µL of LB and grown overnight at 37˚C. 200 µL of these cells were then seeded into 2 L LB with 50 µg mL^{-1} kanamycin and 50 µg mL^{-1} carbenicillin. The culture was grown at 30°C to an OD_{600}~0.6, cooled on ice with swirling to 16°C and induced with 0.5 mM IPTG. Induction continued for 16 hours at 12°C. Cells were harvested by centrifuging for 15 minutes (3,300 RCF at 4°C) and resuspended in 20 mL of lysis buffer (30 mM HEPES [pH 7.5], 0.25 mM EDTA, 5% (v/v) glycerol). The resuspended cell paste was either frozen in liquid nitrogen and stored at -80°C, or prepared for lysis by adding 250 mM NaCl, 1 mM phenylmethanesulfonyl fluoride (PMSF, Sigma-Aldrich), and 1x HALT protease inhibitor (Thermo-Fisher). Cells were lysed in a homogenizer (Avestin), and centrifuged (140,000 RCF, at 4°C) for 35 minutes. A home-made 7 mL SP FF column (resin from GE Healthcare) was equilibrated with 35 mL of buffer A (30 mM HEPES [pH 7.5], 0.25 mM EDTA, 5% (v/v) glycerol) + 200 mM NaCl and the clarified lysate was loaded onto the equilibrated column at 0.5 mL min^{-1}. The column was then washed with 35 mL of buffer A + 250 mM NaCl and the protein was eluted with a 100 mL gradient to buffer B (30 mM HEPES [pH 7.5], 0.25 mM EDTA, 5% (v/v) glycerol, 1 M NaCl) at 1 mL min^{-1}. yRFC-containing fractions were pooled and developed through a 1 mL Q HP column (GE Healthcare) pre-equilibrated with buffer A + 100 mM NaCl. The combined sample from the SP column was loaded onto the Q column at a rate of 0.7 mL min^{-1}. The column was washed with 10 mL of buffer A + 110 mM NaCl at 0.8 mL min^{-1}, and then eluted with an 18 mL gradient

into buffer B at 0.8 mL min^{-1}. The eluted Q column fractions were analyzed on a 10% SDS-PAGE gel, and the fractions containing yRFC were combined, frozen in liquid nitrogen, and stored at -80°C. The yRFC concentration was determined by comparison to a BSA titration curve using SDS-PAGE.

Table 4.1. Oligonucleotides used in this study

Name	Sequence
AD006	**TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT**TCTTCCCATGGTGCGATCGCTCTTCG
YK105	/5Phos/TGCATGCGGCCGCTCTTCC**TCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG**ATGGTGCGATCGCTCTTCG
MB032	/5Phos/TGCATGCGGCCGCTCTTCCCATGGTGCGATCGCTCTTCG
AD012	AGTCTGGATAGCCATAAGTG
AD013	GTAACCACATACTTCCTGCC
AD016	GCAGTCTGTCAGTCAGTGCG
AD017	CGAGGGCATTGCAGTAATTG
AD022	CGTTCATGGCTGAACTCCTG
AD023	CGGCATAACATGCAGTGGAC
AD024	GTGATGTTGCTGCGCTCGATG
AD025	CTCAGCCTGGGTCATTGAAG
AD027	GCTACCACCATGACTAACGC
AD028	GGATATCAGAGCTATGGCTC
AD031	CCGCGATTGCAGATGTTATC
AD032	CTATACAGCCAAGCTTGCAG
IF001	/5Phos/GGGCGGCGACCT/3BioTEG
IF002	/5Phos/AGGTCGCCGCCC/3Dig_N
IF003	/5Phos/AGGTCGCCGCCC/3BioTEG
IF004	/5Phos/GGGCGGCGACCT/3Dig_N
YK_PCNA01	/5phos/GATGACGACAAGCATCATCATCAT

Table 4.1: continued.

YK_PCNA02	GTCTTTGTAGTCGCTGCTGCTGCCCATGGTAT

** The nucleotides for 5'-ssDNA flap or $(CAG)_{13}$ are shown in **bold**.
** The complementary nucleotides to lambda are shown in blue.

Results

yRFC loads yPCNA on λ-DNA containing various extrahelical DNA structures

To investigate *S. cerevisiae* PCNA (yPCNA) dynamics on modified λ-DNA containing a abnormal DNA structure such as Flap or TNR-mediated step loop, I tested how yRFC loads yPCNA on various DNA structures. RFC loads PCNA upon ATP hydrolysis by opening and closing the clamp ring. Although RFC preferentially loads PCNA on primer-template junctions, it can also load on nonspecific homoduplex dsDNA. RFC has also been proposed to load PCNA on extrahelical triplet nucleotide repeats (TNR), ultimately driving TNR expansion. However, most previous biochemical studies have used a truncated RFCΔN that lacks a DNA-binding patch in Rfc1 subunit(232). Because the N-terminal region of Rfc1 is responsible for binding dsDNA non-specifically, people have avoided the motif complicating ensemble experiments. Thus, I investigated yPCNA loading with the more physiologically relevant full-length (wild-type; wt) yRFC.

To test the relative specificity of yPCNA loading onto DNA containing a 30-nt 5'-ssDNA flap, a $(CAG)_{13}$ repeat, and homoduplex dsDNA, I first inserted each of the structure at site A on λ-DNA (Figure 4.2A). For visualization of yPCNA, I purified yPCNA with a triple FLAG epitope tag on the N-terminus, which is far away from the DNA binding site(233, 234). To capture the yRFC-yPCNA complex at the loading site, I used the slowly

hydrolyzing ATPγS in the pre-incubation and loading reactions for the two proteins. ATPγS prevents RFC from closing and releasing PCNA ring, so I can observe the binding sites of PCNA trapped by RFC. yPCNA was fluorescently labeled in situ with α-FLAG Qdots, as described for the single-molecule study of human PCNA (235). I titrated the yRFC(ATPγS)-yPCNA complex concentration to see about one PCNA molecule per DNA. I confirmed that the yPCNA loading on DNA was RFC- and the nucleotide- dependent.

I visualized that yRFC preferentially loads yPCNA at the 5'-ssDNA flap relative to homoduplex regions on DNA, which agrees well with prior studies (Figure 4.2B) (235, 236). Turning off buffer flow retracted both the DNA and yPCNA to the Cr barrier, confirming that yPCNA was bound on the DNA (Figure 4.2B, bottom). I also observed mild enrichment of yPCNA at a (CAG)13 repeat relative to homoduplex DNA (Figure 4.2C). To better characterize the binding efficiency on these structures, I measured the relative enrichment of yPCNA in a 5 kb window spanning the replaced region (site A) for the three recombinant DNA substrates (Figure 4.2C). The window was selected because it captures ~ 99% (~3 standard deviations of the Gaussian fit in Figure 4.2C, top) of all site-specifically bound yPCNA molecules. Loading at the 5'-ssDNA flap shows 2.7-fold enrichment over homoduplex DNA, whereas loading at the (CAG)13 was 1.5-fold higher than homoduplex DNA (Figure 4.2D).

RFC loads PCNA preferentially at a trinucleotide repeat structure

Because the enrichment at $(CAG)_{13}$ was very modest, I also tested the yPCNA binding to $(CAG)_{13}$ inserted at site B and at site A with a flipped tethering geometry (Figure 4.3). The distribution of yPCNA loading on the DNA showed mild preference at the

expected $(CAG)_{13}$ locus in all three DNA substrates (Figure 4.3B). Moreover, the enrichment was statistically significant (p-value: 1 x 10-4) when compared against all possible 5 kb windows across the entire 48.5 kb DNA (Figure 4.4). These data demonstrate that wt yRFC has a mild preference for yPCNA loading onto $(CAG)_{13}$ repeat relative to homoduplex DNA.

PCNA freely diffuses on DNA predominantly by 1D sliding

PCNA has a 3.4 nm-diameter inner opening that is larger than the width of B-form DNA (233, 234). This inner opening facilitates sliding past small mismatches and bulges, but may be blocked by larger extrahelical structures (236). To visually test this hypothesis, I used double-tethered DNA curtains to observe yPCNA diffusion on various DNA substrate (Figure 4.5A). For these experiments, I used ATP to load yPCNA by yRFC so that RFC can hydrolyze ATP to catalyze the closing of the yPCNA ring and release of itself from DNA. To ensure complete yRFC removal from DNA, I washed the flowcells with BSA buffer containing 300 mM NaCl. Following yRFC removal, the yPCNA molecules freely diffused over the entire length of the double-tethered DNA curtains (Figure 4.5).

Ring-like proteins can diffuse on DNA via a combination of two different modes: 1D sliding and/ or hopping. During 1D sliding, the protein tracks the helical backbone of the dsDNA. In contrast, hopping is characterized by a series of correlated microscopic dissociation and re-association events. To differentiate between sliding and hopping, I measured yPCNA diffusion coefficients at increasing ionic strengths (Figure 4.6). A higher ionic strength increases electrostatic screening between a protein and DNA, reducing the fraction of time that a protein is in contact with the DNA. This results in increased diffusion

coefficients at higher ionic strengths. yPCNA diffusion was weakly dependent on the ionic strength, in agreement with a previous study with hPCNA (235). The ionic strength-dependent change in diffusion coefficients was 0.33 ± 0.04 μm^2 sec^{-1} mM^{-1} and 0.25 ± 0.06 μm^2 sec^{-1} mM^{-1} for hPCNA and yPCNA, respectively (Figure 4.6). This indicates that yPCNA as well as hPCNA diffuses predominantly by 1D sliding along the helical pitch of the DNA backbone.

PCNA diffusion is largely blocked by extrahelical structures on λ-DNA

I next monitored yPCNA diffusion on DNA containing a 5'-ssDNA flap or a $(CAG)_{13}$ repeat (Figure 4.6). yPCNA can approach the secondary structure from one of two orientations: (i) in the *cosL* → *cosR* direction, yPCNA encounters a 4-nt ssDNA gap prior to the flap structure and (ii) in the *cosR* → *cosL* direction, yPCNA encounters the 5'-T_{30} tail. I observed that out of 40 yPCNA molecules, 90% (N=36/40) were blocked by the flap in either orientation (N=18 for *cosL* → *cosR* and N=18 for *cosR* → *cosL*). The remaining molecules were either captured at the nicing cassette (5%, N=2/40) or directly bypassed (5%, N=2/40) the flap structure. This observation may result from incomplete re-ligation of the oligos annealed at this site. These data show that the flap is too large to be accommodated within the yPCNA ring. The $(CAG)_{13}$ repeat also blocked yPCNA diffusion from either direction. Remarkably, I also observed that 30% (N=11/36) of the yPCNA molecules loaded directly at the $(CAG)_{13}$ were stationary (Figure 4.6C). To further verify whether $(CAG)_{13}$ structure actually capture yPCNA, I tested the DNA harboring the same TNR repeat at site B as well as flipped DNA with $(CAG)_{13}$ at site A (Figure 4.7A-B). The distribution of diffusing, stationary, and $(CAG)_{13}$-blocked yPCNAs was statistically

indistinguishable across all three $(CAG)_{13}$-containing DNA (Figure 4.7C). I conclude that yPCNA diffusion is blocked by large extrahelical structures and that yRFC loads yPCNA on the $(CAG)_{13}$ repeat. This result is consistent with a mechanism where TNR-bound PCNA interacts with DNA mismatch repair factors to promote TNR expansion (130).

Discussion

Using nickase-based strategy for incorporating extrahelical structures at defined positions along the DNA duplex, I demonstrated that wtRFC preferentially loads yPCNA on 5'-ssDNA flaps and $(CAG)_{13}$ repeats relative to homoduplex DNA. yPCNA can diffuse on homoduplex DNA but is blocked by both a $(CAG)_{13}$ repeat and a 30-nt $(T)_{30}$ ssDNA flap. These observations suggest that PCNA cannot simultaneously accommodate both the homoduplex DNA track (2 nm diameter for B-form) and the additional DNA structures within its inner ring. Molecular dynamics simulations and ensemble biochemical studies suggest that homopolymeric poly-T oligonucleotides occupy an ensemble of extended structures (e.g., 2.7 nm end-to-end distance for a $(T)_{12}$ oligonucleotide; we used a $(T)_{30}$)(237). Similarly, structural and biochemical studies have established that relatively short $(CNG)_{n=3-10}$ repeats are in a hairpin-stem structure with mismatched A•A base pairs that are flanked by two Watson-Crick G•C base pairs (238–241). However, above a critical threshold of trinucleotide repeats, $(CNG)_{n>11}$, the oligonucleotide transitions from a simple hairpin–stem model to bis- (or multi-branched) hairpin-like structures and bulged out ssDNA loops (239, 241). This dynamic, multi-branched structure likely explains why a

$(CAG)_{13}$ repeat acts as both a loading site and a barrier to PCNA diffusion. Finally, yPCNA that is directly loaded onto the $(CAG)_{13}$ is trapped at the lesion. These results shed light on PCNA diffusion dynamics and further highlight the utility of this molecular toolkit for both single-molecule and ensemble biochemical studies.

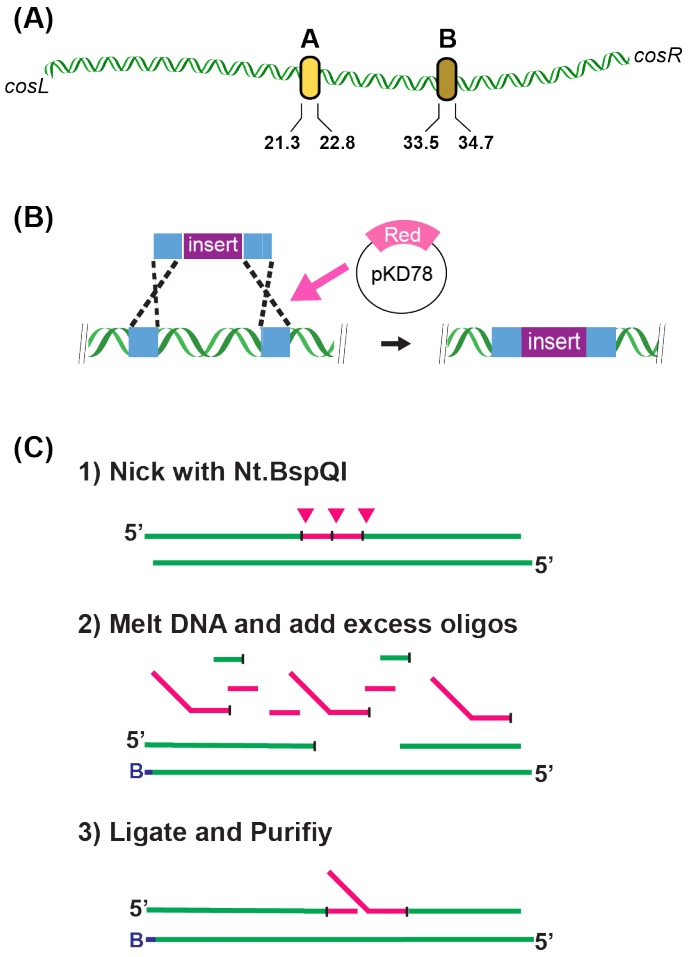

Figure 4.1: **Modification of λ-DNA at unique positions**. (A) Two disposable segments were chosen to modify the λ-DNA. These regions are 21.3 kb and 33.5 kb away from *cosL* (designated A and B respectively). (B) Schematic of the Red-based recombineering. A designed insert DNA containing appropriate homology replaced the original λ-DNA via Red products. (C) Schematic of nicking enzyme-mediated insertion of a synthetic oligonucleotide.

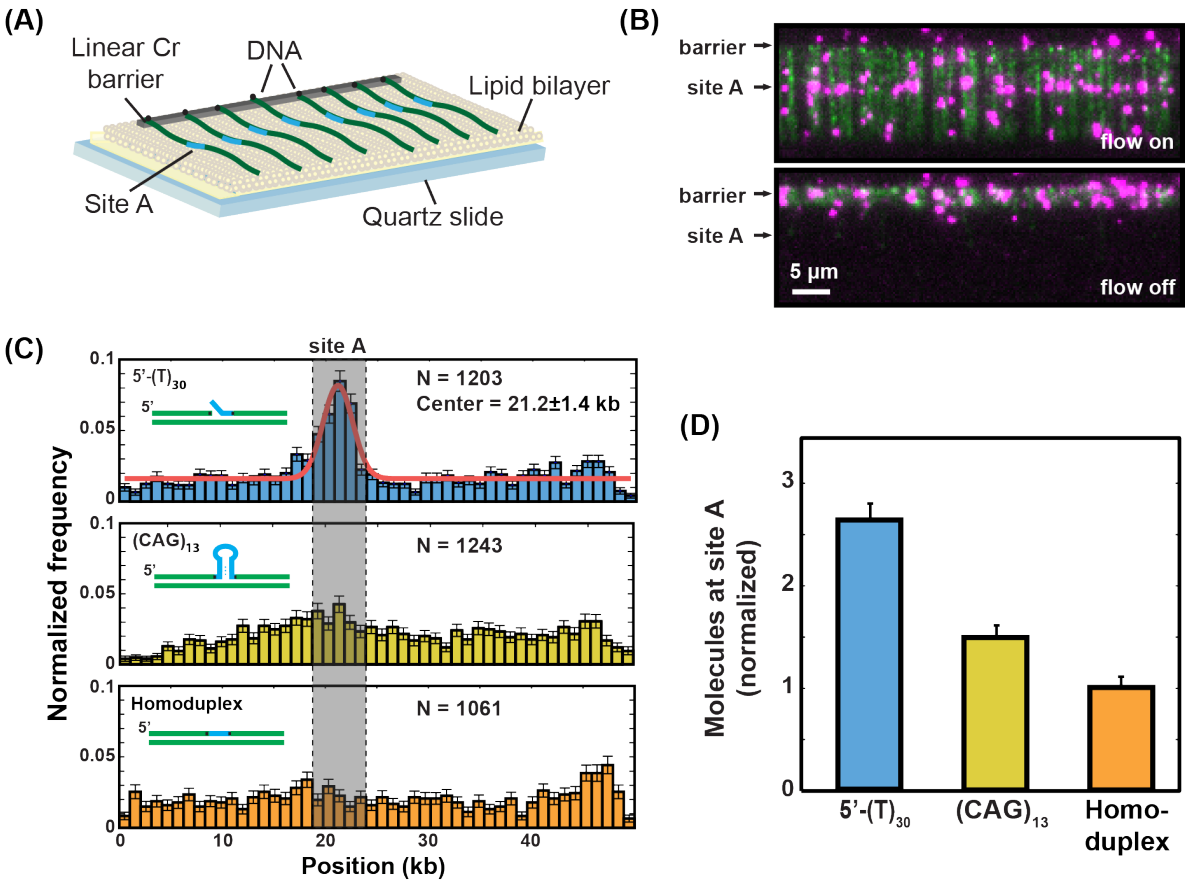

***Figure 4.2*: yRFC loads yPCNA onto various extrahelical structures**. (A) Schematic of a single-tethered DNA curtain assay. DNA molecules (green) are stretched by buffer flow. One of the insertion sites (site A) is indicated in blue. (B) Fluorescent images of yPCNA on λ-DNA containing a 5'-ssDNA flap inserted at site A with buffer flow (top) and without flow (bottom). DNA is stained with YOYO-1 (green) and yPCNA is labeled with anti-FLAG antibody and QDs (magenta). Stopping buffer flow retracts both yPCNA and DNA to the Cr barrier, confirming that yPCNA is bound on the DNA, not on the surface (bottom panel). (C) Binding distribution histogram of yPCNA/yRFC on λ-DNA containing a 5'-ssDNA flap (top), (CAG)13 (middle), and homoduplex (bottom) at site A. Red line is the fit to a Gaussian distribution (center = 21.2 kb ± 1.4 kb). Gray box indicates a 5 kb window that captures 99% of all yPCNAs at site A. (D) The mean number of yPCNA molecules loaded on the site A (within the gray box). Error bars is generated by the bootstrap analysis(242).

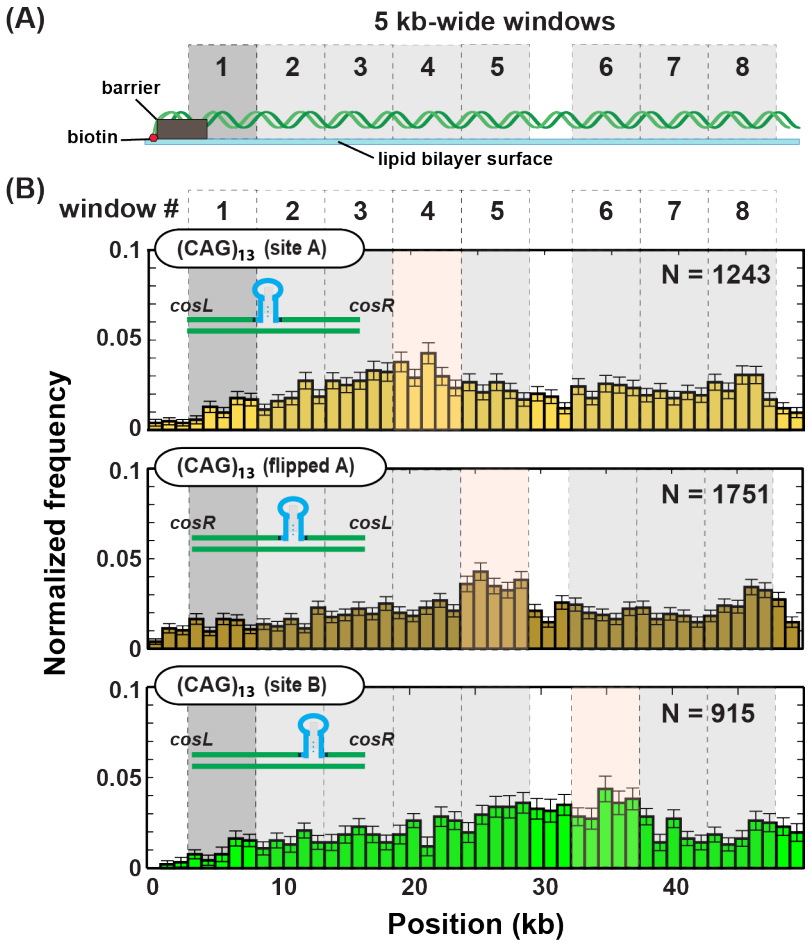

***Figure 4.3*: yRFC has a mild preference for yPCNA loading at $(CAG)_{13}$ repeats relative to homoduplex dsDNA regions.** (A) To determine whether the mild yPCNA loading at $(CAG)_{13}$ structure is authentic observation, the λ-DNA substrate was divided into eight 5 kb-wide windows (dashed lines). The first window (dark gray) is partially obstructed by the Cr barriers. Windows containing the $(CAG)_{13}$ repeat are shown in pink. (B) Normalized binding distribution of yPCNA/yRFC complexes on three DNA substrates containing $(CAG)_{13}$. The binding histograms were also divided into eight 5 kb-wide windows. The first window partially obscured by the Cr barrier is shown in dark gray, the target-containing window is pink, and other homoduplex DNA regions are gray. The (CAG)13-containing windows (pink) show mild preference for PCNA loading for all three substrates.

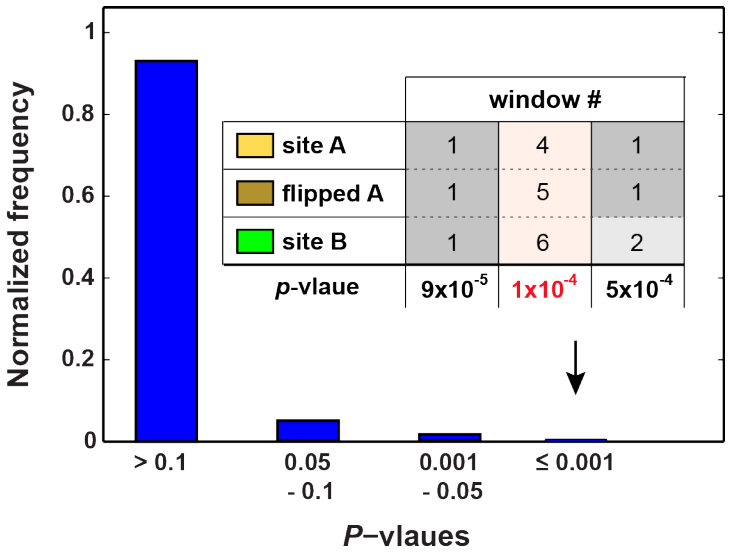

***Figure 4.4*: Statistical analysis of the mild preference for yPCNA loading at $(CAG)_{13}$ repeat on the three DNA substrates**. Based on the data shown in Figure 4.3, A two-tailed t-test was used to compare the groupings of three windows—including the three $(CAG)_{13}$-containing window—against all homoduplex DNA windows. A normalized histogram of all two-tailed t-tests comparing the mean yPCNA occupancy of all three-window combinations relative to the mean yPCNA occupancy in all windows containing homoduplex DNA. Over 97% of the tests showed no significance (p-value > 0.05). The highest p-values are shown in the inset and include the $(CAG)_{13}$-containing sites, as well as the partially obstructed first window. This is because the Cr barrier causes the first window to underestimates yPCNA binding. All p-values in the p=0.001-0.05 range were from window groupings that included two of three $(CAG)_{13}$ repeat windows or window #1.

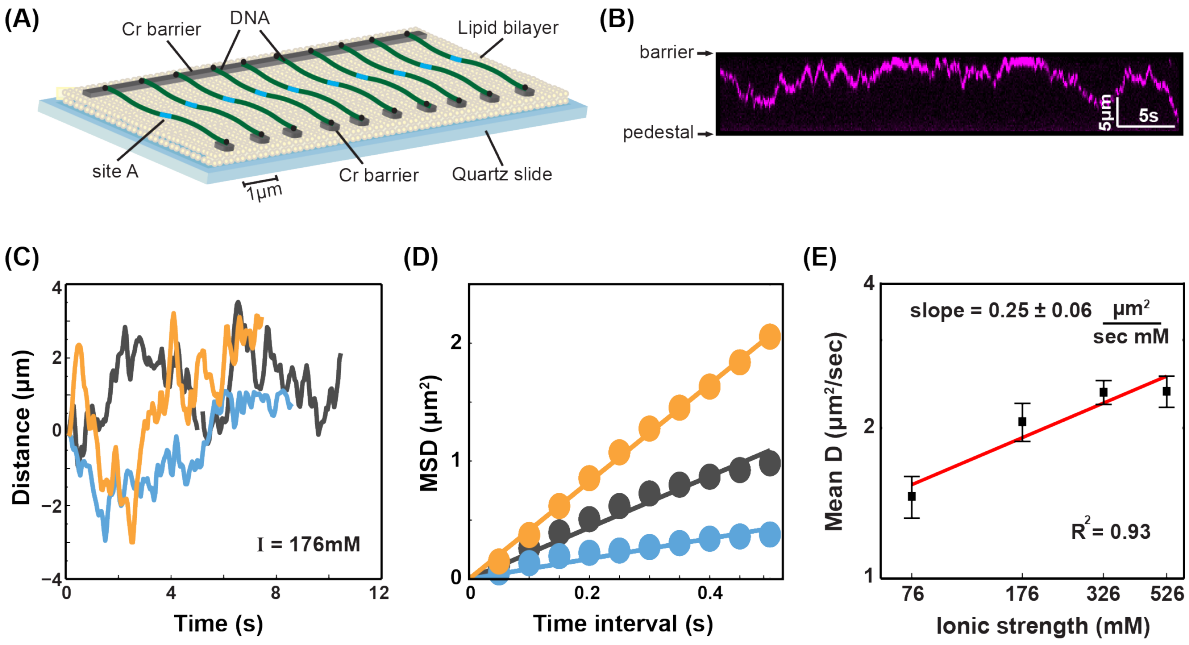

***Figure 4.5*: Characterization of yPCNA diffusion on homoduplex λ-DNA.** (A) Schematic of double-tethered DNA curtains for observing PCNA diffusion on DNA. (B) A kymograph of yPCNA freely diffusing on λ-DNA. (C) Representative single-molecule traces of the position of individual yPCNA molecules (shown in three different colors) on double-tethered DNA curtains at a total ionic strength I = 176 mM. (D) Mean squared displacement (MSD) of each of the molecules in (C). The MSDs are fit to a line, and the slopes are used to calculate the one-dimensional diffusion coefficients (solid lines). (E) Mean yPCNA diffusion coefficients as a function of the ionic strengths (error bars: S.E.M; to N = 30, 29, 29, 31 for 76 mM, 176 mM, 326 mM, and 525 mM ionic strengths, respectively). The red line indicates a linear fit through the data with a slope of 0.25 ± 0.06 μm^2 $(sec\ mM)^{-1}$. The error in the slope represents the standard error of the fit. These results are consistent with a single-molecule study that looked at hPCNA diffusion on homoduplex DNA (243). Diffusion of both hPCNA and yPCNA was weakly dependent on the ionic strength (0.33 ± 0.04 μm^2 sec^{-1} mM^{-1} and 0.25 ± 0.06 μm^2 sec^{-1} mM^{-1} for hPCNA and yPCNA, respectively).

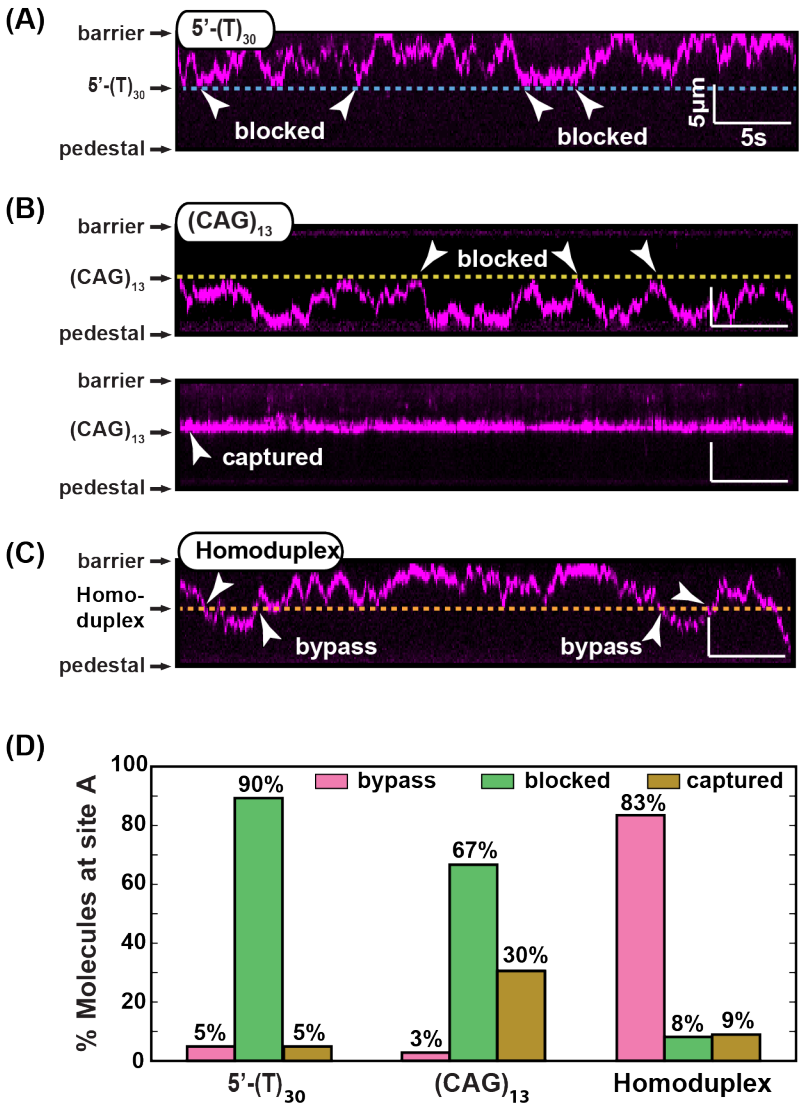

***Figure 4.6*: PCNA diffusion on λ-DNA containing various extrahelical structures.** Kymograph of diffusing PCNA molecule on DNA substrates with (A) a 5'-ssDNA flap, (B) a $(CAG)_{13}$ repeat, or (C) mock-treated homoduplex DNA. The characteristic changes in diffusion behavior of PCNA are indicated with arrows, and the dashed lines indicate site A. (D) Percentage of molecules showing either bypass, blocked, or captured behavior at site A. At least 35 DNA molecules were analyzed and classified into each of three categories (N=40, 36, and 79 for the flap, $(CAG)_{13}$, and mock DNA substrates).

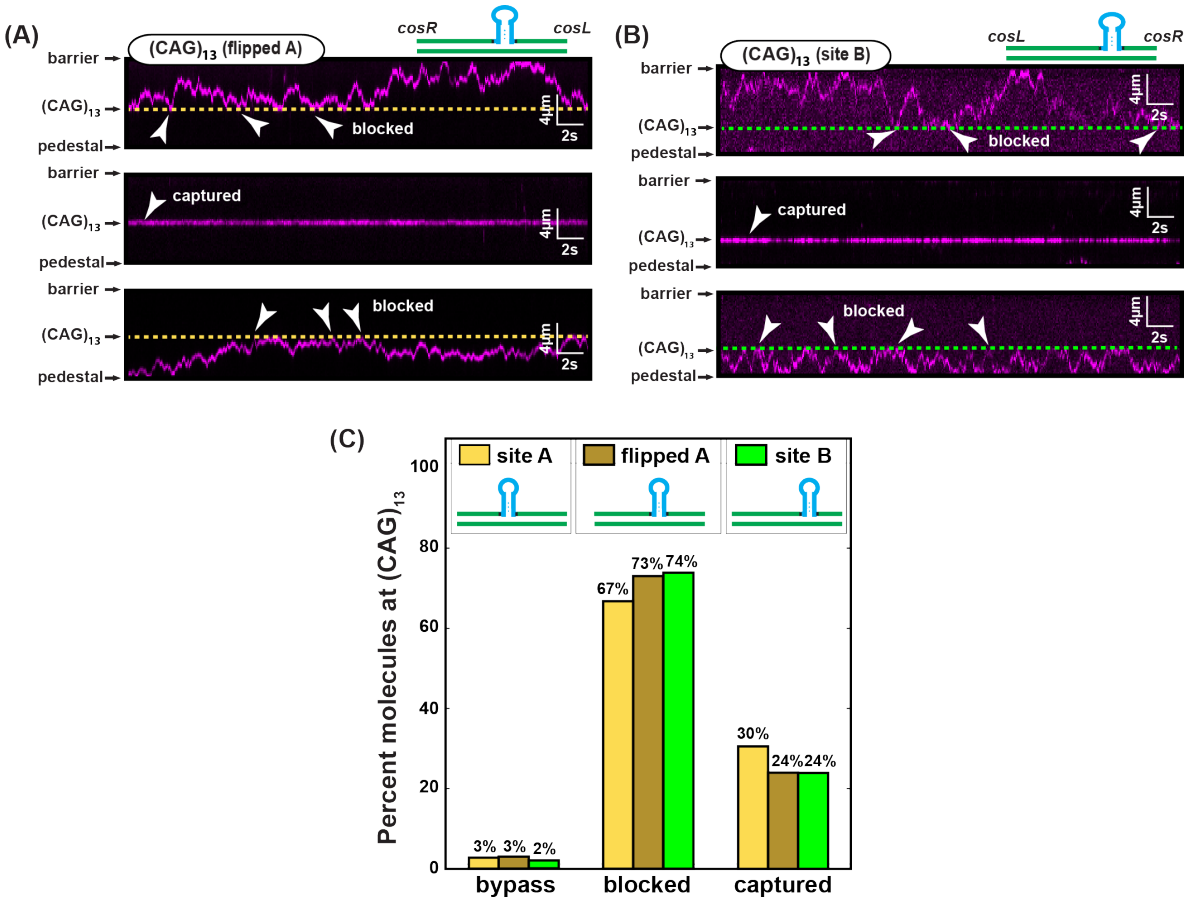

Figure 4.7: yPCNA diffusion on λ-DNA containing $(CAG)_{13}$ repeat at site A and B. Representative kymographs of diffusing yPCNA molecules on DNA having $(CAG)_{13}$ at flipped A (A) or site B (B). The characteristic changes in diffusion behavior of yPCNA at $(CAG)_{13}$ sites are shown with arrows, and each dashed line indicates the position of a $(CAG)_{13}$ repeat. (C) Percentage of molecules showing either bypass, blocked, or captured behavior at $(CAG)_{13}$ insert on λ-DNA. At least 35 DNA molecules were analyzed and classified into each of three categories (N=36, 62, and 46 for site A (from Figure 4B), flipped A, and site B).

Chapter 5. DNA Condensation by Transcription Factors

Introduction

Prion-like domains (PrLD) are a subset of intrinsically disordered regions (IDR) that are characterized by low sequence complexity such as glutamine/asparagine repeats (244). Such sequence features are found in many prions, and the web-based PrLD prediction algorithm, PLAAC, is routinely used to identify PrLDs by these sequence features (244, 245). Approximately 1.2% of protein-coding genes are comprised of PrLD, and about 20% of PrLD-containing DNA binding proteins have known or predicted transcription factor activity (244, 245). Many PrLD-containing TFs act as hubs in regulatory protein networks and can interact with multiple partners via PrLDs (4, 12, 110, 111). Importantly, the sequence properties of PrLDs in non-prion proteins confer the ability to self-associate and form higher-order assemblies such as gel-like structures (246). The functional role of PrLD in protein assembly has been recently described in many cellular pathways, including RNA-binding proteins, membrane receptors, and chaperone proteins (21, 244, 247, 248).

Super-enhancers (SE) are a cluster of enhancers that drive robust gene expressions controlling cell identity and development (116, 249). Recent studies have shown that a high-density of transcriptional coactivators, Mediator, and RNA polymerase II are all located at the enhancer domain (112, 250). Interestingly, IDRs concentrated the transcription apparatus in cells and in nuclear extract (112). IDRs provide the biophysical property necessary to induce molecules to separate into dense and dilute phases (112). This phase separation behavior of transcriptional regulators containing IDRs suggested a model

where transcription machinery locally concentrate all cis- and trans-regulatory elements and facilitate gene transcription activity at SEs (112, 115). However, the phase separation of transcription apparatus on DNA has not been demonstrated at the molecular level. Furthermore, transcription factors (TF) containing PrLDs, a subset of IDRs, have not been explored in this context. Here, I study the role of PrLDs in TF phase separation on DNA using the TF network controlling cell identity in *C. albicans*.

C. albicans is a fungal pathogen that reversibly switches between two cell types, 'white' and 'opaque' (Figure 5.1A). White cells are more virulent in systemic infections whereas opaque cells are the mating-competent form of *C. albicans* (251). Each type can be inherited and maintained across many generations (118). Efg1 is a master transcription factor that is critical for switching from the opaque to white cell type (Figure 5.1B). Efg1 is also important in transcriptional networks controlling hyphal and biofilm formation, which is required for virulence and disease progression (121, 123, 124). Subsequent studies have identified other TFs in the white-opaque circuit, including Wor1, Wor4, and Czf1 (118, 120, 252). Although it is established that these master TFs create a network to control the cell type switching, the mechanism by which these proteins act together is unclear.

Efg1 contains a helix-loop-helix motif that is important for sequence-specific DNA binding (Figure 5.1C) (121, 253). Efg1 also contains extensive polyglutamine (polyQ) stretches at the N- and C-terminal regions (121) that are predicted to be disordered PrLDs by the PLAAC algorithm (Figure 5.1C) (245). My collaborator, Dr. Corey Frazer at Brown University has observed that Efg1 forms liquid-like droplets that undergo fission and fusion at above a threshold concentration *in vitro* (in preparation). Deletion of the PrLD of Efg1

eliminated droplet-like properties, and failed to induce cell type switching (unpublished data). Here, I investigate how the PrLD-containing TF behave on DNA. I demonstrate that i) Efg1 condense DNA in a PrLD- and concentration-dependent manner, ii) Efg1 can compact nucleosome-containing DNA, iii) Efg1 can recruit other master PrLD-containing TFs on DNA independent of their DNA-binding domains. This supports the hypothesis that such master TFs phase separate in a PrLD-dependent manner on DNA and create a condensed transcriptional apparatus that can serve as hub for recruitment of transcriptional coactivators and RNA polymerase II.

Materials and Methods

Purification of Efg1 and TEV Protease: Wild-type Efg1 was overexpressed in BL21 (DE3) cells, and purified from Ni-NTA resin and Superdex-200. All variant Efg1 proteins were provided by the Bennet lab (Brown Univ.). TEV Protease was purified from plasmid pRK793 cotransformed with pRIL (pIF 189) into BL21 (DE3) cells as previously described (Tropea et al., 2009). pRK793 was a gift from David Waugh (Addgene plasmid # 8827).

Imaging DNA condensation by TFs: All single-molecule experiments were conducted in imaging buffer (40 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 0.2 mg mL^{-1} BSA, 50 mM NaCl, 1 mM DTT). DNA contraction was observed via a fluorescent signal on the Dig-labeled DNA ends. These ends were fluorescently labeled by injecting 100 μL of 10 nM α-Dig antibodies (Life Tech, 9H27L19) and 700 μL of 2 nM α-rabbit antibody-conjugated quantum dots (QDs) (Life Tech, Q-11461MP) into the flowcell. After labeling dig-ends of

DNA, the single-tethered DNA molecules were elongated by consistently applying 450 μL min^{-1} flow rate. To see TFs-driven DNA condensation, 10 – 300 nM of the indicated TF was incubated with 100 μg $μL^{-1}$ of TEV protease in 1 mL imaging buffer for 5 minutes at room temperature, then injected into the flowcell at a flow rate of 450 μL min^{-1}. The position of QD-labeled DNA ends was recorded for up to 20 minutes.

Observing TF recruitment via the prion-like domains: Double-tethered DNA curtains were used to determine whether TFs can interact via their prion-like domains. In this assay, the DNA is captured and extended between a chromium barrier and an α-Dig antibody deposited on a chromium pedestal(254). Keeping the DNA fully extended prevents TF-driven compaction. Next, 300 nM of 6xHis–MBP–Efg1 was first injected without TEV cleavage, then 300 nM GFP-Efg1ΔDBD or GFP-Wor1ΔDBD incubated with 100 μg/μL TEV for 5 minutes was injected onto the Efg1-coated DNA molecules.

Results

Efg1 condenses DNA in a concentration-dependent manner

Based on our observation that Efg1 phase separates *in vitro* and forms liquid-like droplets, I reasoned that self-assembly of DNA-bound TFs will lead to contraction of extended DNA molecules. I used high-throughput DNA curtains to test this hypothesis on single molecules of DNA (Figure 5.3A). To visualize TF-induced DNA contraction, I used the single-tethered DNA curtains where the DNA was biotinylated on one end and labeled with a fluorescent antibody on the opposite end (Figure 5.3A). Efg1 was used as a

representative TF in most DNA curtain experiments. To test whether Efg1 condenses DNA, I first pre-incubated the TF with TEV protease to remove the MBP and injected the mixture into the flowcell containing pre-assembled DNA curtains. Efg1 rapidly contracted the DNA as observed by the retraction of the fluorescent DNA end towards the biotinylated tether point (Figure 5.3B, top). I could measure the rate and length of DNA condensation by using particle tracking of the fluorescent DNA ends. The rate of DNA contraction was Efg1 concentration-dependent, with no contraction at 10 nM Efg1 (N=27 DNA molecules) and rapid contraction of 4 ± 1.4 kb s^{-1} at 300 nM Efg1 (Figure 5.3C). Efg1 contracted most DNA molecules nearly completely (to the microfabricated barrier) when using both the high (300 nM) and intermediate (50 nM) Efg1 concentration (Figure 5.3D). In contrast, MBP-Efg1 (without treatment with TEV protease) significantly slowed the DNA contraction rate and reduced the overall contraction length, indicating that the MBP largely blocks Efg1 self-assembly on DNA (Figure 5.3B-D). This is consistent with our collaborator's observation that MBP-Efg1 does not phase separate *in vitro* (unpublished data).

Prion-like domains are required for DNA condensation

Next, I tested whether the DNA condensation requires both the Efg1 DNA binding domain as well as the PrLD. For this purpose, I used two GFP-tagged Efg1 constructs that delete either the DNA binding domain (DBD) or the PrLD. As expected, injecting GFP-Efg1 that lacked the PrLD (GFP-Efg1ΔPrLD) still retained robust DNA-binding activity and could completely coat the DNA substrate (Figure 5.4A, top and middle). Also, injecting GFP-Efg1 that lacked the DBD neither condense nor bound the DNA (Figure

5.4A, bottom). These data demonstrate that the Efg1-driven DNA condensation requires both the PrLD self-assembly and DBD domains.

I next sought to determine whether the PrLDs are sufficient for multivalent TF interactions on DNA. In this assay, the DNA molecules were tethered by both DNA ends to inhibit Efg1-driven DNA condensation (Figure 5.4B). Next, Efg1 was injected into the flowcell, followed by GFP-TF fusions that lack DNA-binding domains (GFP-Efg1ΔDBD or GFP-Wor1ΔDBD). Both GFP-Efg1ΔDBD and GFP-Wor1ΔDBD rapidly accumulated over the entire length of the double-tethered DNA molecules, indicating that these proteins were interacting with DNA-bound Efg1 via their PrLDs (Figure 5.4C).

Nucleosome regulates DNA condensation by the transcription factors

In the cell, TFs must function in the context of chromatin. Thus, I assessed how nucleosomes regulate Efg1-induced DNA condensation (Figure 5.5A). For this assay, dense nucleosome arrays (>10 nucleosomes per DNA) were deposited on the DNA substrate via step-wise salt dialysis (132). Nucleosomes were visualized via a fluorescent antibody directed at an HA epitope on H2A, as described previously(50, 206). Efg1 still contracted the nucleosomal DNA substrates, but at a significantly slower rate than on naked DNA (Figure 5.5B–D). Interestingly, Efg1 contraction rates changed on single molecules and had frequent intermediate plateaus (Figure 5.5B). I thus reported the average contraction rate per molecule in Figure 5.5C. I reasoned that the reduced compaction rates may stem from nucleosomes acting as physical barriers for Efg1 binding. In support of this model, nucleosome-free DNA regions compacted more rapidly than nucleosome-dense regions of the same DNA substrate (arrows in Figure 5.5B). Thus, nucleosome density can

control local DNA compaction by TFs *in vivo*. Taken together, these data demonstrate that Efg1 condenses naked and nucleosome-coated DNA and that multiple TFs can be recruited to the same DNA substrate via interactions between the PrLDs.

Discussion

SEs create a large enhancer domain mediated by a high-density of transcriptional regulators. This proposes a model where phase-separated transcriptional regulators form a concentrated apparatus to control gene expression (112, 115, 257). The phase-separated transcriptional coactivators and Mediator have been observed *in vivo*, and colocalize with RNA polymerase II (112, 250). However, no direct evidence of phase separation of TFs on DNA at a molecular level has been reported so far. Here, I used a single-molecule assay and a TF network controlling white-opaque switching in *C. albicans* as a case study. I demonstrated that master TFs containing PrLDs self-associate on DNA, which results in DNA condensation. The DNA compaction by Efg1, a TF responsible for white cell type and virulence in *C. albicans*, required both DNA binding and the PrLD. The presence of an MBP tag next to the N-terminal PrLD of Efg1 significantly inhibited the DNA condensation. These results support a model by which the multivalent interactions between PrLDs facilitate the phase separation of TFs on DNA and create the DNA-TFs condensates. I also show that DNA covered by one master TF can be also bound by other master TFs through PrLDs. I propose that the condensates made by one or multiple master TFs serve as a platform that recruits other multivalent coactivators and RNA polymerase II to create a bona fide active enhancer domain (258, 259).

DNA is organized into chromatin, which influences transcriptional condensation by TFs *in vivo*. My single-molecule experiments with nucleosomal DNA show that the TF can condense DNA with nucleosomes, albeit at a slower rate and a lesser extent of condensation (Figure 5.4B–D). The DNA region with dense nucleosomes was condensed after the sparse nucleosomes region (Figure 5.4B). This is consistent with previous reports that nucleosomes represent a mechanical barrier to regulate gene expression, and that regulatory sequences including promoters are nucleosome-depleted to promote TF accessibility across eukaryotic species (260–262). Thus, I propose that nucleosomes are physical barriers for TF binding at SEs, and nucleosome-density regulates transcription activation by controlling the DNA-TFs condensation and/or condensate dissolution *in vivo*.

Efg1 is also an important factor in hyphal growth for biofilm formation and virulence by *C. albicans* (263, 264). The TFs controlling biofilm development in the species create an additional TF network containing Efg1 and factors different from those regulating white-opaque switching (263). Using PLAAC analysis, I found that most reported TFs in the biofilm network contain PrLDs including Tec1, Bcr1, Brg1, Not80. How does Efg1 distinctively interact with each networking PrLD-containing TFs? Whether these TFs condense nucleosomal DNA and form distinct condensates from the white-opaque switching network are all interesting areas of future study with implications for how PrLDs modulate complex TF networks.

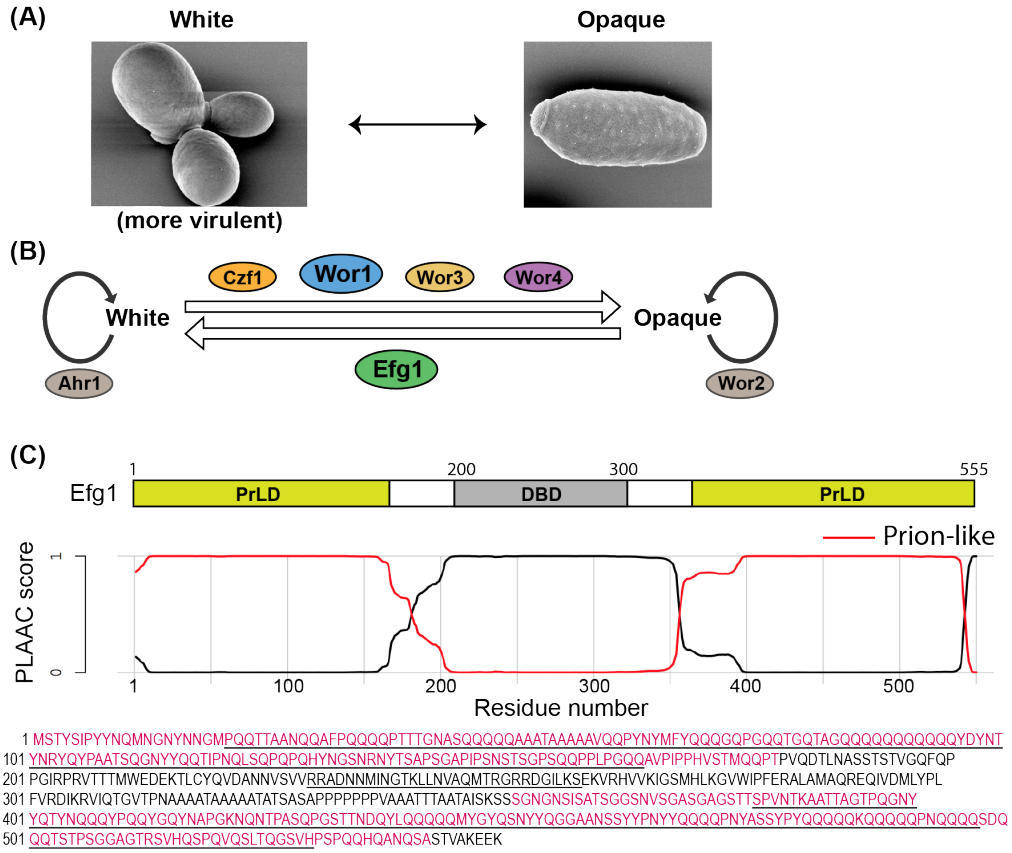

Figure 5.1: **Efg1, a major transcription factor in white-opaque switching network, has long PrLDs at N- and C-terminal regions.** (A) Representative microscope images of *C. albicans* in the white (left) and opaque (right) cell type. (obtained from Bennett Lab in Brown Univ.) (B) A schematic model of transcriptional network controlling the white-opaque cell type switching in *C. albicans* (122, 124). Efg1 is required to switch from opaque to and maintain white cell type (121). (C) Predicted PrLDs using PLAAC analysis (245). DNA-binding domain (DBD) is shown based on previous functional mapping of the Efg1 (121, 253). The locations of PrLDs are indicated in red in the graph and one-letter amino acid residues of Efg1. This analysis also predicted IDRs by FoldIndex (265), which is underlined in sequence residues (black underline).

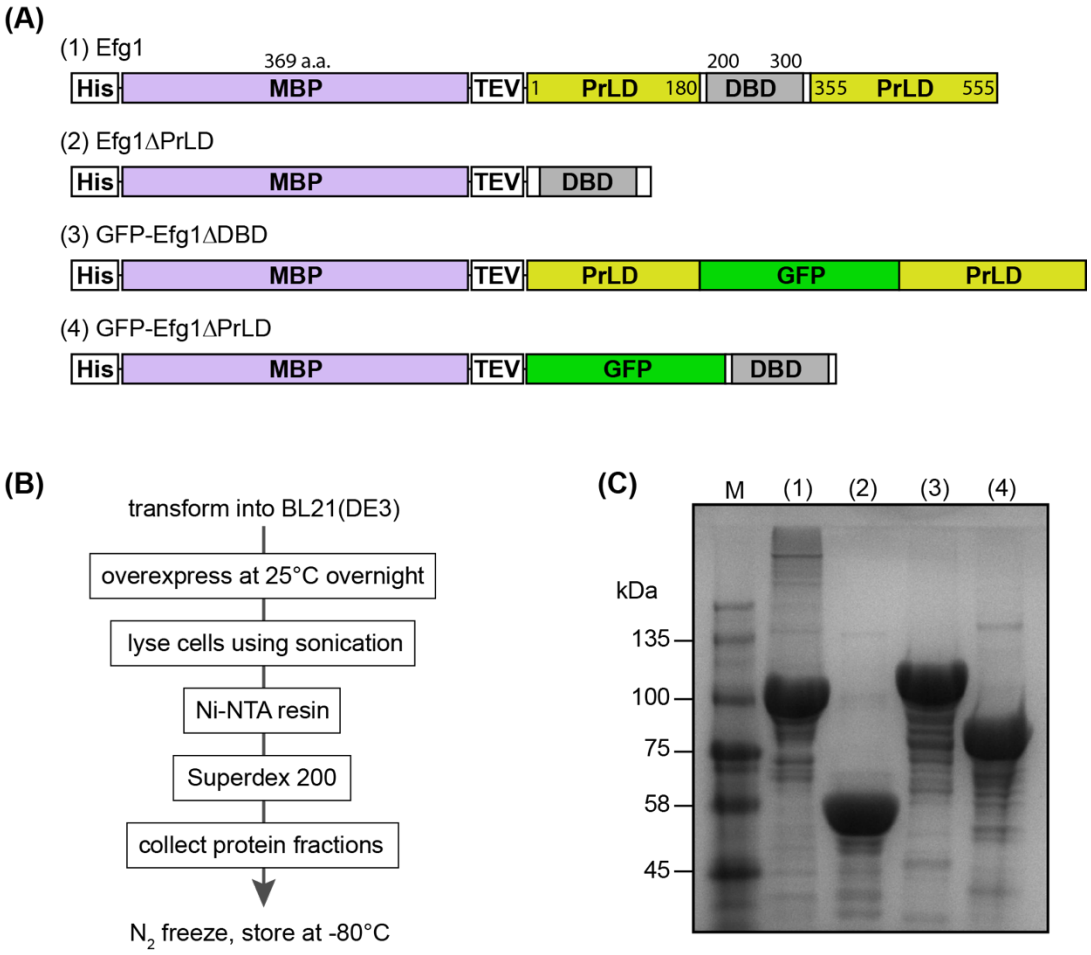

***Figure 5.2*: Illustration of Efg1 constructs**. (A) Wild-type Efg1 and other variants were purified with 6xHis- and MBP-tags at N-terminus of Efg1. GFP was used to visualize the variants in some assays. TEV cleavage site was added between the tags and the proteins of interest. (B) Purification scheme of Efg1 constructs. (C) 8% SDS-PAGE gel of all 4 purified constructs (M indicates protein ladder).

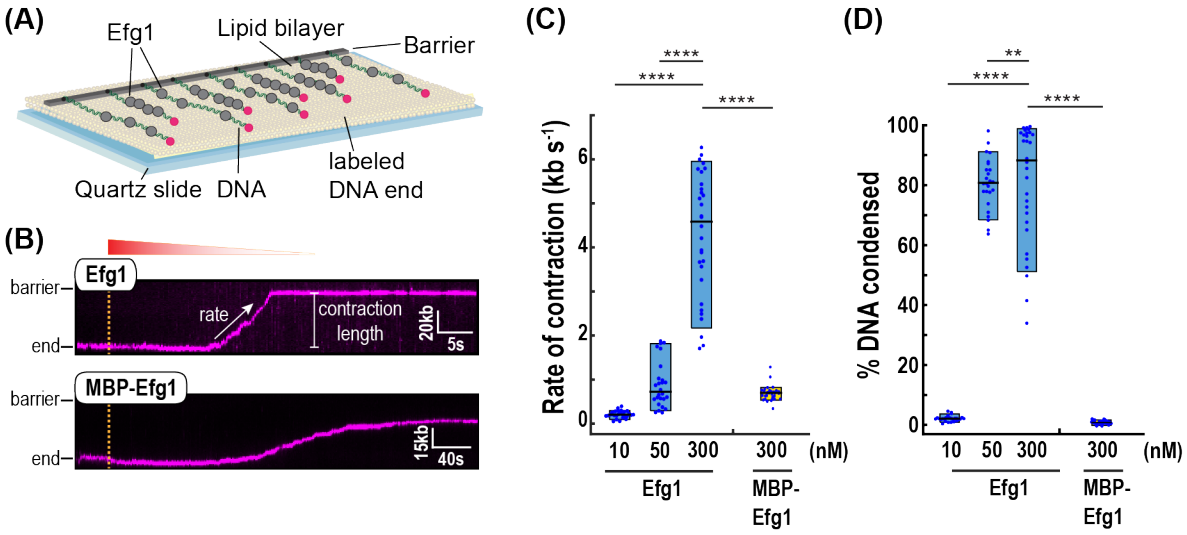

Figure 5.3: **Efg1 condenses DNA in a TEV-cleavage and concentration dependent manner.** (A) Schematic of the DNA curtains assay. DNA ends are fluorescently labeled with Qdot-conjugated α-Dig antibodies and dark Efg1 is injected into the flowcell while keeping the DNA extended via mild buffer flow. (B) A representative kymograph of 300 nM Efg1 condensing naked DNA in the presence (top) and absence (bottom) of TEV protease. The time point when Efg1 is injected into the flowcell is indicated with yellow dashed lines. (C) Rate and (D) degree of DNA condensation expressed as a percent of the total DNA length in the presence and absence of nucleosomes at different Efg1 concentrations. Boxplots indicate the median, 10th, and 90th percentiles of the distribution. P-values are obtained from K-S test: * *P*-values <0.05, ** *P*-value < 0.01, *** *P*-value < 0.005, and **** *P*-value < 0.0001. N>25 DNA molecules for all experiments.

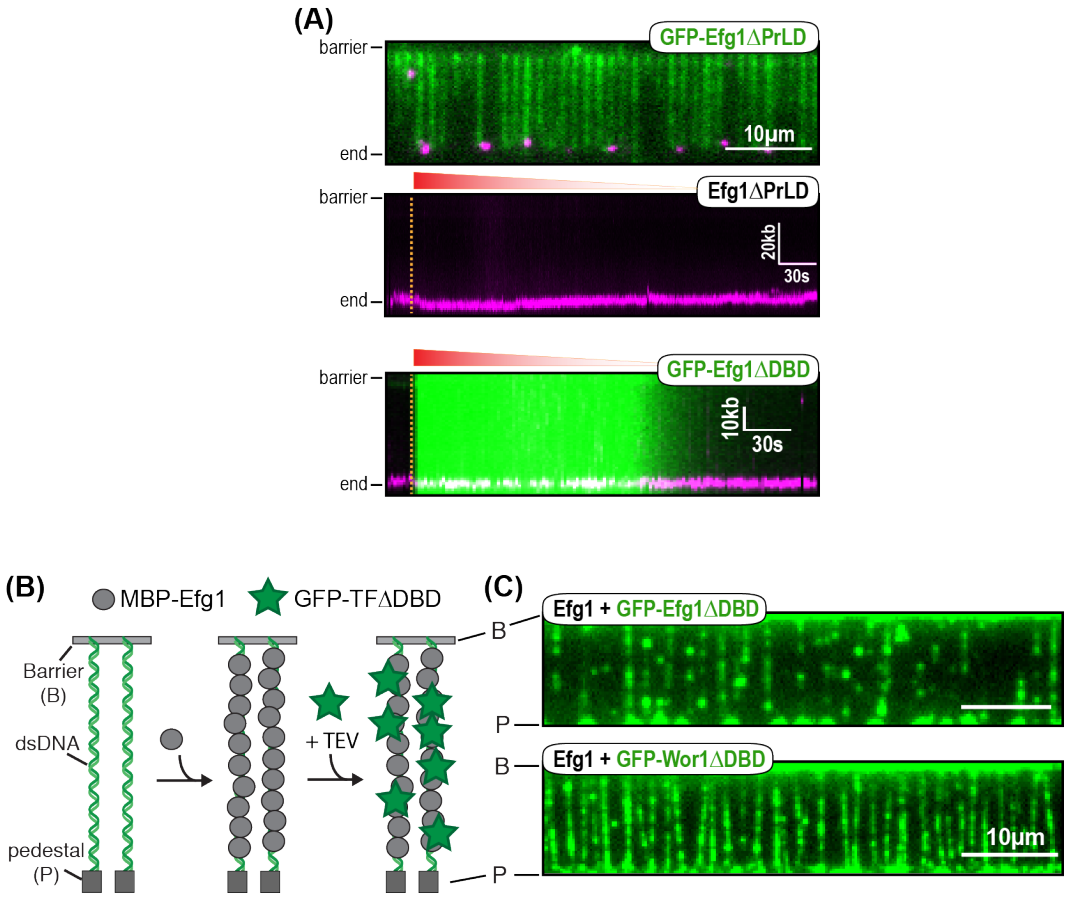

Figure 5.4: **PrLD is required for the multivalent interaction of TFs.** (A) top: An image of single-tethered DNA curtain after injecting 300 nM GFP-Efg1ΔPrLD into the flowcell; A representative kymograph showing nearly stationary DNA ends (magenta) when injecting 300 nM Efg1ΔPrLD (middle) or GFP-Efg1ΔDBD (green, bottom). (B) An illustration and (C) fluorescent images demonstrating that DNA-bound Efg1 can recruit other TFs via prion-like domains. (B) DNA molecules are double-tethered to block the TF-driven DNA condensation, and 300 nM wild-type (wt) Efg1 was first incubated with the DNA. Then, GFP-Efg1ΔDBD or GFP-Wor1ΔDBD was injected with TEV protease. (C) Kymographs showing recruitment of GFP-Efg1ΔDBD (top) or GFP-Wor1ΔDBD (bottom) to DNA-bound Efg1.

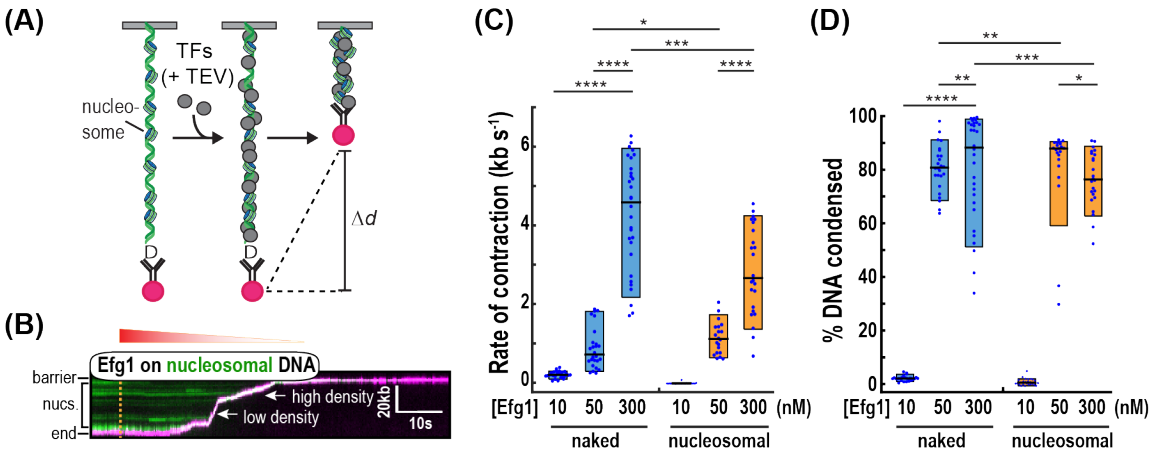

Figure 5.5: **Nucleosome regulates DNA condensation by Efg1.** (A) Schematic of nucleosome-DNA condensation by TFs. (B) A kymograph of Efg1 condensing nucleosome-coated DNA. Nucleosomes are shown in green and the fluorescently labeled DNA end is in magenta. The time point when Efg1 is injected into the flowcell is indicated with yellow dashed lines. The protein traverses the flowcell for a few minutes as its concentration is diluted by constant buffer flow. The rate and extent of DNA condensation is measured by tracking the fluorescent DNA end. (C) Rate and (D) degree of naked and nucleosome DNA condensation expressed as a percent of the total DNA length in the presence and absence of nucleosomes at different Efg1 concentrations. Boxplots indicate the median, 10th, and 90th percentiles of the distribution. P-values are obtained from K-S test: * *P*-values <0.05, ** *P*-value < 0.01, *** *P*-value < 0.005, and **** *P*-value < 0.0001. N>25 DNA molecules for all experiments.

Chapter 6. Summary and Future Directions

In this dissertation, I described how IDRs are critical for the proper regulation of DNA repair and transcription-associated proteins. To observe dynamic behavior of these proteins in a more physiological *in vitro* setting, I developed and characterized DNA curtains assembled with human nucleosomes (50) (Chapter 2). In Chapter 3, I discussed the detailed functions of IDRs in Mlh1-Pms1, a eukaryotic MMR protein. MMR proteins also have been implicated in TNR expansions in human neurodegenerative diseases (61, 71). In Chapter 4, I also showed that RFC loads PCNA, another essential MMR protein, on the extrahelical structure made by TNR sequences (130). Lastly, I described in Chapter 5 that IDRs in transcription factors lead to DNA condensation and that nucleosomes regulate this activity, which suggests the mechanism of SE formation.

IDRs in Mlh1-Pms1 are essential for efficient MMR *in vivo* (109). There are a few points to think about the mechanistic role of the IDR linkers. First, Mlh1-Pms1 IDRs could impact the interaction with other target proteins such as ExoI and PCNA. Mlh1 C-terminal domain has a site for interaction with the Mlh1 interacting protein box ((R/K)SK(Y/F)F motif) of ExoI, which is critical for ExoI-dependent MMR (266, 267). Also, a direct interaction between PCNA and Mlh1-Pms1 has been reported by gel filtration (268). Further structural studies are required to elucidate how the linkers are connected to the binding sites in Mlh1-Pms1 and how the binding of partners is influenced by the ATP-induced change in linker conformations. Second, the IDRs also affect the Mlh1-Pms1 endonuclease activity. Work described in this thesis shows that the IDRs also influenced

the endonuclease activity of Mlh1-Pms1 (Figure 3.6). Notably, the single-round nicking activity on circular DNA was unaffected while the multiple-round nickng was compromised for the DLD_{null}. The nicking activity requires ATP as well as PCNA (Figure 3.6A). Direct AFM imaging of Mlh1-Pms1 showed ATP-dependent compaction of the complex and suggested a model of linker scrunching upon nucleotide binding (97). Based on these observations, the IDR linkers could mediate the allosteric regulation of the nucleolytic activity by ATP binding. The compromised nicking activity for DLD_{null} could be due to the altered ATP-induced linker rearrangement. This can be examined by directly observing the linker conformation using AFM or EM imaging. Alternatively, one can conduct FRET single-molecule experiments to determine the dynamic conformational movement of the IDR linkers with respect to the NTD ATPase and endonuclease domain of Mlh1-Pms1. Third, our data have suggested the role of Mlh1-Pms1 IDRs in facilitating diffusion on DNA (Figure 3.4). 1D diffusion of DNA-binding proteins is affected by several factors such as the ionic strength, DNA binding affinity, and friction between DNA and a moving protein (44, 53, 269). DNA binding affinity of DLD_{null} were unaffected (Figure 3.2) whereas the diffusion coefficients were significantly decreased compared to wild-type (Figure 3.3). Mlh1-Pms1 is thought to encircle DNA by binding through a positively-charged groove made under the dimerized N-terminal domains. Thus, I speculate that shortening the linkers could bring the N- and C-terminal domains closer and thereby lead to an increase in the translational friction between the protein and DNA increase.

The IDR linkers of MutL homologs vary in length. For example, the prokaryotic MutL linkers (90 a.a.) are only half as long as that in Mlh1 (109, 270). Interestingly, the portion of disordered residues tends to be high in eukaryotic homologs and the duplicated gene Pms1 (Pms2 in human) (Figure 3.1). Moreover, the Mlh3 homolog also has a long IDR linker (117 a.a.) (189). Besides its roles in MMR, Mlh3 plays additional roles in meiotic recombination, and the Mlh1-Mlh3 complex also prefers specific DNA structures (189, 271). These observations are consistent with the proteomic analysis showing that IDRs are more common in eukaryotes than prokaryotes and that they tend to be prevalent in regulatory proteins that require sophisticated roles in complex protein networks (1, 2). It will be interesting to see how Mlh1-Mlh3 IDRs play roles in MMR and meiotic recombination.

MMR proteins function on chromatin, where nucleosomes are assembled. Recent studies have shown that the presence of nucleosomes compromises Msh2-6 binding and sliding activity, and histone chaperones such as CAF-1 regulate MMR efficiency *in vitro* (37, 272, 273). The question thus arises: how do MMR proteins locate their targets in the presence of nucleosomes *in vivo*? Guo-Min Li's group suggested a model by which human Msh2-6 is recruited to chromatin before replication by directly binding trimethylated histone 3 at H3K36 (H3K36me3) via a Pro-Trp-Trp-Pro motif in the Msh6 domain (274, 275). Nucleosomes are disassembled ahead of replication fork and the nucleosome-bound Msh2-6 is released to the free DNA in search of a lesion. This motif is not present in yeast Msh6. However, yeast Msh2-6 can be recruited to the replication fork through a PCNA-interacting motif (PIP-box) (78, 276). Importantly, *in vivo* cell imaging with fluorescent

proteins in yeast revealed that Mlh1-Pms1 foci are formed later than Msh2-6 foci that colocalized with replication factories (78). Thus, it is likely that Mlh1-Pms1 find their target after nucleosomes are re-assembled behind the replication fork. I showed that Mlh1-Pms1 loads onto nucleosomal DNA nonspecifically and bypasses nucleosomes during 1D diffusion (Chapter 3) (163, 277). I speculate that Mlh1-Pms1 is loaded on a DNA region distant from a mismatch. Mlh1-Pms1 efficiently bypasses the nucleosome barriers newly deposited behind the replication machinery (163, 277). Further study is required to see if other chromatin-associated factors such as histone modifications or remodelers promote Mlh1-Pms1 recruitment on chromatin.

Nucleosomes can serve as a beneficial controller for MMR. In a defined system that included both Mlh1-Pms1 and ExoI, nucleosomes could slightly promote MMR efficiency by inhibiting excess degradation of the discontinuous strand (272). I reason that the limited degradation in the presence of nucleosomes resulted from the combination of two mechanisms: i) nucleosomes blocked the degradation by ExoI (165) and ii) the reduced nucleosome bypass activity of Mlh1-Pms1 upon ATP binding (Figure 3.5) restrained the excision distance. How nucleosomes differentially influence the degradation of lesion-containing DNA in ExoI-dependent and ExoI-independent pathway needs further investigation.

Msh2-3 differs from Msh2-6 with regards to its specificity for lesions, its target search mechanism, and its binding to PCNA and Mlh1-Pms1. However, how Msh2-3-dependent MMR is differentially regulated on chromatin remains unclear. Which factor(s) influences Msh2-3 recruitment on chromatin and how does Msh2-3 regulate other MMR

factors? The physical interaction of Msh2-3 with PCNA through PIP-box is required for MMR, but the mechanism by which the two proteins influence MMR efficiency is not clear (278). PCNA acts as a signal for strand discrimination and activates Mlh1-Pms1 nicking activity, which is required for efficient MMR (279, 280). As shown in Msh2-6, Msh2-3 might also inhibit PCNA unloading via PIP-box and maintain strand-specific MMR capability (280). Interestingly, unlike Msh2-6, Msh2-3 interacts with Mlh1-Pms1 through the same site that is required to bind PCNA (281). This suggests that Msh2-3-initiated MMR is fundamentally different from the Msh2-6-dependent pathway. It will be interesting to see how the competing Msh2-3 interaction between PCNA and Mlh1-Pms1 influences the PCNA roles in strand-specific MMR and Mlh1-Pms1 activation.

MMR factors have been implicated in TNR instability that leads to many neurodegenerative diseases (62, 63, 71, 282). The length of TNR sequences is found to be expanded in the affected patients (61, 63). Expanded TNR in human disease form hairpin structure *in vitro*, which is thought to be bound by MMR factors and cause replication stalling (282, 283). In Chapter 4, I demonstrated the preferential loading of PCNA on $(CAG)_{13}$ structure by RFC (Figure 4.2 and Figure 4.3) (225). The $(CAG)_{13}$ repeat captured the loaded PCNA, preventing 1D diffusion by the sliding clamp (Figure 4.6). I propose the two following mechanisms of MMR-mediated TNR instability. First, many PCNA molecules can be loaded. The redundantly loaded PCNA molecules may recruit the partners including Msh2-3 and Mlh1-Pms1, leading to the formation of a large protein complex. The mass of MMR factors on a TNR site will result in replication stalling and slippage of the daughter strand over its template, leading to TNR instability (283). Another

possibility is that the captured PCNA on a TNR structure no longer plays a role in strand discrimination in the absence of DNA discontinuities (202, 280). Then, PCNA loading on a TNR may not direct Mlh1-Pms1 endonuclease on the TNR-containing strand and thus fail to remove the DNA lesion. This will allow DNA polymerase to use the TNR sequences as a template over and over, which can cause TNR expansion. It will be interesting to see how the PCNA loading on TNR-containing substrate impair the strand-specific MMR capability and how such defects influence TNR instability.

MMR proteins are also required for class switch recombination (CSR) of antibody genes in B cells (284–286). In the model of MMR-mediated CSR, Msh2-6 is critical for recognizing the U:G mismatches created by AID, a cytosine deaminating enzyme. The U:G mismatch bound by Msh2-6 recruits downstream factors such as PCNA, Mlh1-Pms2, and ExoI (284, 287). Interestingly, the MMR proteins promote double-strand breaks (DSB) at the switch repeat regions of the Ig genes. The ATPase and endonuclease activity of Mlh1-Pms2 seem to play critical roles in DSB formation during CSR (286). Moreover, the ATPase activity of Mlh1-Pms2 influences DNA end processing and triggers non-homologous end joining (NHEJ) by recruiting other interacting partners such as DNA-PKcs (286, 288). Further studies are needed to understand the mechanisms by which Mlh1-Pms2 promotes CSR in both upstream and downstream of DSB and its role in favoring NHEJ.

In Chapter 5, I have shown that IDR-containing TFs condense DNA molecules into puncta and discussed the role of nucleosome barriers in active enhancer domains for gene transcription. The observations that TFs mediate DNA condensation via PrLD interactions

opens many questions in the phase separation and SE fields. For example, how does the transcriptional machinery perform active transcription on DNA condensed by these TFs *in vivo*? A recent study has suggested that heterochromatin protein 1 (HP1) forms phase-separated droplets and compact heterochromatin DNA into puncta, suggesting a role of the HP1-driven condensates in gene silencing (289). Then, what factors make the transcriptional apparatus dynamic and accessible during transcription by RNA polymerase II in euchromatin? If nucleosomes regulate function, how is nucleosome density determined and maintained so that it makes a local DNA dense enough for close proximity among regulatory sequences but also makes the region loose enough for efficient gene activation by a transcription machinery? It is well established that histone acetylation loosens condensed chromatin and specific histone marks, such as acetylation of histone H3 lysine 27 (H3K27ac) or H3 lysine 56 (H3K56ac), are enriched in active enhancers (290, 291). It would be interesting to see how such histone acetylation may either promote or inhibit TFs-mediated nucleosome DNA condensation.

References

1. Babu,M.M. (2016) The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society Transactions*, **44**, 1185–1200.

2. van der Lee,R., Buljan,M., Lang,B., Weatheritt,R.J., Daughdrill,G.W., Dunker,A.K., Fuxreiter,M., Gough,J., Gsponer,J., Jones,D.T., *et al.* (2014) Classification of intrinsically disordered regions and proteins. *Chem. Rev.*, **114**, 6589–6631.

3. Patil,A. and Nakamura,H. (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett.*, **580**, 2041–2045.

4. Haynes,C., Oldfield,C.J., Ji,F., Klitgord,N., Cusick,M.E., Radivojac,P., Uversky,V.N., Vidal,M. and Iakoucheva,L.M. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput. Biol.*, **2**, e100.

5. Sherry,K.P., Johnson,S.E., Hatem,C.L., Majumdar,A. and Barrick,D. (2015) Effects of Linker Length and Transient Secondary Structure Elements in the Intrinsically Disordered Notch RAM Region on Notch Signaling. *Journal of Molecular Biology*, **427**, 3587–3597.

6. Uversky,V.N., Oldfield,C.J. and Dunker,A.K. (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.*, **18**, 343–384.

7. Wright,P.E. and Dyson,H.J. (2014) Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology*, **16**, 18–29.

8. Anthis,N.J. and Clore,G.M. (2013) The Length of the Calmodulin Linker Determines the Extent of Transient Interdomain Association and Target Affinity. *Journal of the American Chemical Society*, **135**, 9648–9651.

9. Walsh,M.P. (1983) Calmodulin and its roles in skeletal muscle function. *Can Anaesth Soc J*, **30**, 390–398.

10. Wojciak,J.M., Martinez-Yamout,M.A., Dyson,H.J. and Wright,P.E. (2009) Structural basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation domains. *EMBO J.*, **28**, 948–958.

11. Kasper,L.H., Fukuyama,T., Biesen,M.A., Boussouar,F., Tong,C., de Pauw,A., Murray,P.J., van Deursen,J.M.A. and Brindle,P.K. (2006) Conditional knockout mice reveal distinct functions for the global transcriptional coactivators CBP and p300 in T-cell development. *Mol. Cell. Biol.*, **26**, 789–809.

12. Dyson,H.J. and Wright,P.E. (2016) Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300. *J. Biol. Chem.*, **291**, 6714–6722.

13. Disfani,F.M., Hsu,W.-L., Mizianty,M.J., Oldfield,C.J., Xue,B., Dunker,A.K., Uversky,V.N. and Kurgan,L. (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics*, **28**, i75-83.

14. Bergeron-Sandoval,L.-P., Safaee,N. and Michnick,S.W. (2016) Mechanisms and Consequences of Macromolecular Phase Separation. *Cell*, **165**, 1067–1079.

15. Brady,J.P., Farber,P.J., Sekhar,A., Lin,Y.-H., Huang,R., Bah,A., Nott,T.J., Chan,H.S., Baldwin,A.J., Forman-Kay,J.D., *et al.* (2017) Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences*, **114**, E8194–E8203.

16. Banani,S.F., Lee,H.O., Hyman,A.A. and Rosen,M.K. (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.*, **18**, 285–298.

17. Jonas,S. and Izaurralde,E. (2013) The role of disordered protein regions in the assembly of decapping complexes and RNP granules. *Genes & Development*, **27**, 2628–2641.

18. Mitrea,D.M. and Kriwacki,R.W. (2016) Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling*, **14**.

19. Brangwynne,C.P., Eckmann,C.R., Courson,D.S., Rybarska,A., Hoege,C., Gharakhani,J., Julicher,F. and Hyman,A.A. (2009) Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science*, **324**, 1729–1732.

20. Molliex,A., Temirov,J., Lee,J., Coughlin,M., Kanagaraj,A.P., Kim,H.J., Mittag,T. and Taylor,J.P. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell*, **163**, 123–133.

21. Banjade,S. and Rosen,M.K. (2014) Phase transitions of multivalent proteins can promote clustering of membrane receptors. *eLife*, **3**.

22. Li,P., Banjade,S., Cheng,H.-C., Kim,S., Chen,B., Guo,L., Llaguno,M., Hollingsworth,J.V., King,D.S., Banani,S.F., *et al.* (2012) Phase transitions in the assembly of multivalent signalling proteins. *Nature*, **483**, 336–340.

23. Choi,U.B., Kazi,R., Stenzoski,N., Wollmuth,L.P., Uversky,V.N. and Bowen,M.E. (2013) Modulating the intrinsic disorder in the cytoplasmic domain alters the biological activity of the N-methyl-D-aspartate-sensitive glutamate receptor. *J. Biol. Chem.*, **288**, 22506–22515.

24. Stott,K., Watson,M., Howe,F.S., Grossmann,J.G. and Thomas,J.O. (2010) Tail-mediated collapse of HMGB1 is dynamic and occurs via differential binding of the acidic tail to the A and B domains. *J. Mol. Biol.*, **403**, 706–722.

25. Brower-Toland,B., Wacker,D.A., Fulbright,R.M., Lis,J.T., Kraus,W.L. and Wang,M.D. (2005) Specific Contributions of Histone Tails and their Acetylation to the Mechanical Stability of Nucleosomes. *Journal of Molecular Biology*, **346**, 135–146.

26. Bah,A. and Forman-Kay,J.D. (2016) Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *Journal of Biological Chemistry*, **291**, 6696–6705.

27. Erdmann,D., Zimmermann,C., Fontana,P., Hau,J.-C., De Pover,A. and Chène,P. (2010) Simultaneous protein expression and modification: an efficient approach for production of unphosphorylated and biotinylated receptor tyrosine kinases by triple infection in the baculovirus expression system. *J Biomol Tech*, **21**, 9–17.

28. Andres,S.N. and Williams,R.S. (2017) CtIP/Ctp1/Sae2, molecular form fit for function. *DNA Repair*, **56**, 109–117.

29. Warren,C. and Shechter,D. (2017) Fly Fishing for Histones: Catch and Release by Histone Chaperone Intrinsically Disordered Regions and Acidic Stretches. *J. Mol. Biol.*, **429**, 2401–2426.

30. van Leeuwen,H.C. (1997) Linker length and composition influence the flexibility of Oct-1 DNA binding. *The EMBO Journal*, **16**, 2043–2053.

31. Currie,S.L., Lau,D.K.W., Doane,J.J., Whitby,F.G., Okon,M., McIntosh,L.P. and Graves,B.J. (2017) Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4 and ETV5. *Nucleic Acids Research*, **45**, 2223–2241.

32. Vuzman,D. and Levy,Y. (2012) Intrinsically disordered regions as affinity tuners in protein–DNA interactions. *Mol. BioSyst.*, **8**, 47–57.

33. Kozlov,A.G., Weiland,E., Mittal,A., Waldman,V., Antony,E., Fazio,N., Pappu,R.V. and Lohman,T.M. (2015) Intrinsically Disordered C-Terminal Tails of E. coli Single-Stranded DNA Binding Protein Regulate Cooperative Binding to Single-Stranded DNA. *Journal of Molecular Biology*, **427**, 763–774.

34. Kozlov,A.G., Weiland,E., Mittal,A., Waldman,V., Antony,E., Fazio,N., Pappu,R.V. and Lohman,T.M. (2015) Intrinsically disordered C-terminal tails of E. coli single-stranded DNA binding protein regulate cooperative binding to single-stranded DNA. *J. Mol. Biol.*, **427**, 763–774.

35. Kasahara,K., Shiina,M., Higo,J., Ogata,K. and Nakamura,H. (2018) Phosphorylation of an intrinsically disordered region of Ets1 shifts a multi-modal interaction ensemble to an auto-inhibitory state. *Nucleic Acids Research*, **46**, 2243–2251.

36. Oudet,P., Gross-Bellard,M. and Chambon,P. (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, **4**, 281–300.

37. Li,F., Tian,L., Gu,L. and Li,G.-M. (2009) Evidence that nucleosomes inhibit mismatch repair in eukaryotic cells. *J. Biol. Chem.*, **284**, 33056–33061.

38. Adkins,N.L., Niu,H., Sung,P. and Peterson,C.L. (2013) Nucleosome dynamics regulates DNA processing. *Nat. Struct. Mol. Biol.*, **20**, 836–842.

39. Sparmann,A. and van Lohuizen,M. (2006) Polycomb silencers control cell fate, development and cancer. *Nature Reviews Cancer*, **6**, 846–856.

40. Flavahan,W.A., Gaskell,E. and Bernstein,B.E. (2017) Epigenetic plasticity and the hallmarks of cancer. *Science*, **357**, eaal2380.

41. Mirny,L., Slutsky,M., Wunderlich,Z., Tafvizi,A., Leith,J. and Kosmrlj,A. (2009) How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical*, **42**, 434013.

42. Slutsky,M. and Mirny,L.A. (2004) Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.*, **87**, 4021–4035.

43. Riggs,A.D., Bourgeois,S. and Cohn,M. (1970) The lac repressor-operator interaction. 3. Kinetic studies. *J. Mol. Biol.*, **53**, 401–417.

44. Bhattacherjee,A. and Levy,Y. (2014) Search by proteins for their DNA target site: 1. The effect of DNA conformation on protein sliding. *Nucleic Acids Research*, **42**, 12404–12414.

45. Kim,J.G., Takeda,Y., Matthews,B.W. and Anderson,W.F. (1987) Kinetic studies on Cro repressor-operator DNA interaction. *J. Mol. Biol.*, **196**, 149–158.

46. Elf,J., Li,G.-W. and Xie,X.S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, **316**, 1191–1194.

47. Wang,Y.M., Austin,R.H. and Cox,E.C. (2006) Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys. Rev. Lett.*, **97**, 048302.

48. Kochaniak,A.B., Habuchi,S., Loparo,J.J., Chang,D.J., Cimprich,K.A., Walter,J.C. and van Oijen,A.M. (2009) Proliferating cell nuclear antigen uses two distinct modes to move along DNA. *J. Biol. Chem.*, **284**, 17700–17710.

49. Kim,J.H. and Larson,R.G. (2007) Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules. *Nucleic Acids Res.*, **35**, 3848–3858.

50. Brown,M.W., Kim,Y., Williams,G.M., Huck,J.D., Surtees,J.A. and Finkelstein,I.J. (2016) Dynamic DNA binding licenses a repair factor to bypass roadblocks in search of DNA lesions. *Nat. Commun.*, **7**, 10607.

51. Gorman,J., Chowdhury,A., Surtees,J.A., Shimada,J., Reichman,D.R., Alani,E. and Greene,E.C. (2007) Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. *Mol. Cell*, **28**, 359–370.

52. Ramanathan,S.P., van Aelst,K., Sears,A., Peakman,L.J., Diffin,F.M., Szczelkun,M.D. and Seidel,R. (2009) Type III restriction enzymes communicate in 1D without looping between their target sites. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 1748–1753.

53. Givaty,O. and Levy,Y. (2009) Protein Sliding along DNA: Dynamics and Structural Characterization. *Journal of Molecular Biology*, **385**, 1087–1097.

54. Mangel,W.F., McGrath,W.J., Xiong,K., Graziano,V. and Blainey,P.C. (2016) Molecular sled is an eleven-amino acid vehicle facilitating biochemical interactions via sliding components along DNA. *Nature Communications*, **7**, 10202.

55. Kunkel,T.A. (2004) DNA Replication Fidelity. *Journal of Biological Chemistry*, **279**, 16895–16898.

56. Iyer,R.R., Pluciennik,A., Burdett,V. and Modrich,P.L. (2006) DNA Mismatch Repair: Functions and Mechanisms. *Chemical Reviews*, **106**, 302–323.

57. Tachiki,H., Kato,R. and Kuramitsu,S. (2000) DNA binding and protein-protein interaction sites in MutS, a mismatched DNA recognition protein from Thermus thermophilus HB8. *J. Biol. Chem.*, **275**, 40703–40709.

58. Manhart,C.M. and Alani,E. (2017) DNA replication and mismatch repair safeguard against metabolic imbalances. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 5561–5563.

59. Pensotti,V., Radice,P., Presciuttini,S., Calistri,D., Gazzoli,I., Grimalt Perez,A., Mondini,P., Buonsanti,G., Sala,P., Rossetti,C., *et al.* (1997) Mean age of tumor onset in hereditary nonpolyposis colorectal cancer (HNPCC) families correlates with the presence of mutations in DNA mismatch repair genes. *Genes Chromosomes Cancer*, **19**, 135–142.

60. Wanat,J.J., Singh,N. and Alani,E. (2007) The effect of genetic background on the function of Saccharomyces cerevisiae mlh1 alleles that correspond to HNPCC missense mutations. *Hum. Mol. Genet.*, **16**, 445–452.

61. Kovtun,I.V. and McMurray,C.T. (2008) Features of trinucleotide repeat instability in vivo. *Cell Res.*, **18**, 198–213.

62. Budworth,H., Harris,F.R., Williams,P., Lee,D.Y., Holt,A., Pahnke,J., Szczesny,B., Acevedo-Torres,K., Ayala-Peña,S. and McMurray,C.T. (2015) Suppression of Somatic Expansion Delays the Onset of Pathophysiology in a Mouse Model of Huntington's Disease. *PLOS Genetics*, **11**, e1005267.

63. Budworth,H. and McMurray,C.T. (2013) A Brief History of Triplet Repeat Diseases. In Kohwi,Y., McMurray,C.T. (eds), *Trinucleotide Repeat Protocols*. Humana Press, Totowa, NJ, Vol. 1010, pp. 3–17.

64. Panigrahi,G.B., Slean,M.M., Simard,J.P. and Pearson,C.E. (2012) Human mismatch repair protein hMutLα is required to repair short slipped-DNAs of trinucleotide repeats. *J. Biol. Chem.*, **287**, 41844–41850.

65. Bak,S.T., Sakellariou,D. and Pena-Diaz,J. (2014) The dual nature of mismatch repair as antimutator and mutator: for better or for worse. *Front Genet*, **5**, 287.

66. Hsieh,P. and Zhang,Y. (2017) The Devil is in the details for DNA mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 3552–3554.

67. Hou,C., Chan,N.L.S., Gu,L. and Li,G.-M. (2009) Incision-dependent and error-free repair of (CAG)(n)/(CTG)(n) hairpins in human cell extracts. *Nat. Struct. Mol. Biol.*, **16**, 869–875.

68. Bourn,R.L., De Biase,I., Pinto,R.M., Sandi,C., Al-Mahdawi,S., Pook,M.A. and Bidichandani,S.I. (2012) Pms2 Suppresses Large Expansions of the (GAA·TTC)n Sequence in Neuronal Tissues. *PLoS ONE*, **7**, e47085.

69. Gannon,A.-M.M., Frizzell,A., Healy,E. and Lahue,R.S. (2012) MutSβ and histone deacetylase complexes promote expansions of trinucleotide repeats in human cells. *Nucleic Acids Res.*, **40**, 10324–10333.

70. Kantartzis,A., Williams,G.M., Balakrishnan,L., Roberts,R.L., Surtees,J.A. and Bambara,R.A. (2012) Msh2-Msh3 interferes with Okazaki fragment processing to promote trinucleotide repeat expansions. *Cell Rep*, **2**, 216–222.

71. Williams,G.M. and Surtees,J.A. (2015) MSH3 Promotes Dynamic Behavior of Trinucleotide Repeat Tracts In Vivo. *Genetics*, **200**, 737–754.

72. Warren,J.J., Pohlhaus,T.J., Changela,A., Iyer,R.R., Modrich,P.L. and Beese,L.S. (2007) Structure of the Human MutSα DNA Lesion Recognition Complex. *Molecular Cell*, **26**, 579–592.

73. Jónsson,Z.O., Hindges,R. and Hübscher,U. (1998) Regulation of DNA replication and repair proteins through interaction with the front side of proliferating cell nuclear antigen. *EMBO J.*, **17**, 2412–2425.

74. De March,M., Merino,N., Barrera-Vilarmau,S., Crehuet,R., Onesti,S., Blanco,F.J. and De Biasio,A. (2017) Structural basis of human PCNA sliding on DNA. *Nat Commun*, **8**, 13935.

75. Fukui,K. (2010) DNA Mismatch Repair in Eukaryotes and Bacteria. *Journal of Nucleic Acids*, **2010**, 1–16.

76. Owen,B.A.L., H Lang,W. and McMurray,C.T. (2009) The nucleotide binding dynamics of human MSH2-MSH3 are lesion dependent. *Nat. Struct. Mol. Biol.*, **16**, 550–557.

77. Kadyrov,F.A., Holmes,S.F., Arana,M.E., Lukianova,O.A., O'Donnell,M., Kunkel,T.A. and Modrich,P. (2007) Saccharomyces cerevisiae MutLalpha is a mismatch repair endonuclease. *J. Biol. Chem.*, **282**, 37181–37190.

78. Hombauer,H., Campbell,C.S., Smith,C.E., Desai,A. and Kolodner,R.D. (2011) Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates. *Cell*, **147**, 1040–1053.

79. Goellner,E.M., Putnam,C.D. and Kolodner,R.D. (2015) Exonuclease 1-dependent and independent mismatch repair. *DNA Repair*, **32**, 24–32.

80. Hermans,N., Laffeber,C., Cristovão,M., Artola-Borán,M., Mardenborough,Y., Ikpa,P., Jaddoe,A., Winterwerp,H.H.K., Wyman,C., Jiricny,J., *et al.* (2016) Dual daughter strand incision is processive and increases the efficiency of DNA mismatch repair. *Nucleic Acids Res.*, **44**, 6770–6786.

81. Gorman,J., Wang,F., Redding,S., Plys,A.J., Fazio,T., Wind,S., Alani,E.E. and Greene,E.C. (2012) Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E3074-3083.

82. Keijzers,G., Liu,D. and Rasmussen,L.J. (2016) Exonuclease 1 and its versatile roles in DNA repair. *Critical Reviews in Biochemistry and Molecular Biology*, **51**, 440–451.

83. Goellner,E.M., Putnam,C.D., Graham,W.J., Rahal,C.M., Li,B.-Z. and Kolodner,R.D. (2018) Identification of Exo1-Msh2 interaction motifs in DNA mismatch repair and new Msh2-binding partners. *Nat. Struct. Mol. Biol.*, 10.1038/s41594-018-0092-y.

84. Tishkoff,D.X., Boerger,A.L., Bertrand,P., Filosi,N., Gaida,G.M., Kane,M.F. and Kolodner,R.D. (1997) Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 7487–7492.

85. Goellner,E.M., Smith,C.E., Campbell,C.S., Hombauer,H., Desai,A., Putnam,C.D. and Kolodner,R.D. (2014) PCNA and Msh2-Msh6 activate an Mlh1-Pms1 endonuclease pathway required for Exo1-independent mismatch repair. *Mol. Cell*, **55**, 291–304.

86. Smith,C.E., Mendillo,M.L., Bowen,N., Hombauer,H., Campbell,C.S., Desai,A., Putnam,C.D. and Kolodner,R.D. (2013) Dominant Mutations in S. cerevisiae PMS1 Identify the Mlh1-Pms1 Endonuclease Active Site and an Exonuclease 1-Independent Mismatch Repair Pathway. *PLoS Genetics*, **9**, e1003869.

87. Ban,C. and Yang,W. (1998) Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell*, **95**, 541–552.

88. Guarne,A. (2001) Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. *The EMBO Journal*, **20**, 5521–5531.

89. Tannergård,P., Lipford,J.R., Kolodner,R., Frödin,J.E., Nordenskjöld,M. and Lindblom,A. (1995) Mutation screening in the hMLH1 gene in Swedish hereditary nonpolyposis colon cancer families. *Cancer Res.*, **55**, 6092–6096.

90. Shcherbakova,P.V. and Kunkel,T.A. (1999) Mutator phenotypes conferred by MLH1 overexpression and by heterozygosity for mlh1 mutations. *Mol. Cell. Biol.*, **19**, 3177–3183.

91. Räschle,M., Dufner,P., Marra,G. and Jiricny,J. (2002) Mutations within the hMLH1 and hPMS2 Subunits of the Human MutLα Mismatch Repair Factor Affect Its ATPase Activity, but Not Its Ability to Interact with hMutSα. *Journal of Biological Chemistry*, **277**, 21810–21820.

92. Tran,P.T. and Liskay,R.M. (2000) Functional studies on the candidate ATPase domains of Saccharomyces cerevisiae MutLalpha. *Mol. Cell. Biol.*, **20**, 6390–6398.

93. Corbett,K.D. and Berger,J.M. (2005) Structural dissection of ATP turnover in the prototypical GHL ATPase TopoVI. *Structure*, **13**, 873–882.

94. Ban,C., Junop,M. and Yang,W. (1999) Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell*, **97**, 85–97.

95. Yamamoto,T., Iino,H., Kim,K., Kuramitsu,S. and Fukui,K. (2011) Evidence for ATP-dependent structural rearrangement of nuclease catalytic site in DNA mismatch repair endonuclease MutL. *J. Biol. Chem.*, **286**, 42337–42348.

96. Tomer,G. (2002) Contribution of Human Mlh1 and Pms2 ATPase Activities to DNA Mismatch Repair. *Journal of Biological Chemistry*, **277**, 21801–21809.

97. Sacho,E.J., Kadyrov,F.A., Modrich,P., Kunkel,T.A. and Erie,D.A. (2008) Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha. *Mol. Cell*, **29**, 112–121.

98. Bende,S.M. and Grafström,R.H. (1991) The DNA binding properties of the MutL protein isolated from Escherichia coli. *Nucleic Acids Res.*, **19**, 1549–1555.

99. Arana,M.E., Holmes,S.F., Fortune,J.M., Moon,A.F., Pedersen,L.C. and Kunkel,T.A. (2010) Functional residues on the surface of the N-terminal domain of yeast Pms1. *DNA Repair*, **9**, 448–457.

100. Schorzman,A.N., Perera,L., Cutalo-Patterson,J.M., Pedersen,L.C., Pedersen,L.G., Kunkel,T.A. and Tomer,K.B. (2011) Modeling of the DNA-binding site of yeast Pms1 by mass spectrometry. *DNA Repair*, **10**, 454–465.

101. Hall,M.C., Wang,H., Erie,D.A. and Kunkel,T.A. (2001) High affinity cooperative DNA binding by the yeast Mlh1-Pms1 heterodimer. *Journal of Molecular Biology*, **312**, 637–647.

102. Kadyrov,F.A., Dzantiev,L., Constantin,N. and Modrich,P. (2006) Endonucleolytic Function of MutLα in Human Mismatch Repair. *Cell*, **126**, 297–308.

103. Gueneau,E., Dherin,C., Legrand,P., Tellier-Lebegue,C., Gilquin,B., Bonnesoeur,P., Londino,F., Quemener,C., Le Du,M.-H., Márquez,J.A., *et al.* (2013) Structure of the MutLα C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. *Nat. Struct. Mol. Biol.*, **20**, 461–468.

104. Miyaki,M., Nishio,J., Konishi,M., Kikuchi-Yanoshita,R., Tanaka,K., Muraoka,M., Nagato,M., Chong,J.-M., Koike,M., Terada,T., *et al.* (1997) Drastic genetic instability of tumors and normal tissues in Turcot syndrome. *Oncogene*, **15**, 2877–2881.

105. Pillon,M.C., Lorenowicz,J.J., Uckelmann,M., Klocko,A.D., Mitchell,R.R., Chung,Y.S., Modrich,P., Walker,G.C., Simmons,L.A., Friedhoff,P., *et al.* (2010) Structure of the endonuclease domain of MutL: unlicensed to cut. *Mol. Cell*, **39**, 145–151.

106. Iino,H., Kim,K., Shimada,A., Masui,R., Kuramitsu,S. and Fukui,K. (2011) Characterization of C- and N-terminal domains of *Aquifex aeolicus* MutL endonuclease: N-terminal domain stimulates the endonuclease activity of C-terminal domain in a zinc-dependent manner. *Bioscience Reports*, **31**, 309–322.

107. Xue,B., Dunbrack,R.L., Williams,R.W., Dunker,A.K. and Uversky,V.N. (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1804**, 996–1010.

108. Guarné,A., Ramon-Maiques,S., Wolff,E.M., Ghirlando,R., Hu,X., Miller,J.H. and Yang,W. (2004) Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *EMBO J.*, **23**, 4134–4145.

109. Plys,A.J., Rogacheva,M.V., Greene,E.C. and Alani,E. (2012) The unstructured linker arms of Mlh1-Pms1 are important for interactions with DNA during mismatch repair. *J. Mol. Biol.*, **422**, 192–203.

110. Liu,J., Perumal,N.B., Oldfield,C.J., Su,E.W., Uversky,V.N. and Dunker,A.K. (2006) Intrinsic disorder in transcription factors. *Biochemistry*, **45**, 6873–6888.

111. Contreras-Martos,S., Piai,A., Kosol,S., Varadi,M., Bekesi,A., Lebrun,P., Volkov,A.N., Gevaert,K., Pierattelli,R., Felli,I.C., *et al.* (2017) Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. *Scientific Reports*, **7**.

112. Sabari,B.R., Dall'Agnese,A., Boija,A., Klein,I.A., Coffey,E.L., Shrinivas,K., Abraham,B.J., Hannett,N.M., Zamudio,A.V., Manteiga,J.C., *et al.* (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, 10.1126/science.aar3958.

113. Spitz,F. and Furlong,E.E.M. (2012) Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, **13**, 613–626.

114. Whyte,W.A., Orlando,D.A., Hnisz,D., Abraham,B.J., Lin,C.Y., Kagey,M.H., Rahl,P.B., Lee,T.I. and Young,R.A. (2013) Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell*, **153**, 307–319.

115. Hnisz,D., Shrinivas,K., Young,R.A., Chakraborty,A.K. and Sharp,P.A. (2017) A Phase Separation Model for Transcriptional Control. *Cell*, **169**, 13–23.

116. Hnisz,D., Abraham,B.J., Lee,T.I., Lau,A., Saint-André,V., Sigova,A.A., Hoke,H.A. and Young,R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934–947.

117. Odds,F.C. (1988) Candida and candidosis 2. ed. Tindall, London.

118. Lohse,M.B., Ene,I.V., Craik,V.B., Hernday,A.D., Mancera,E., Morschhäuser,J., Bennett,R.J. and Johnson,A.D. (2016) Systematic Genetic Screen for Transcriptional Regulators of the *Candida albicans* White-Opaque Switch. *Genetics*, **203**, 1679–1692.

119. Noble,S.M., Gianetti,B.A. and Witchley,J.N. (2017) Candida albicans cell-type switching and functional plasticity in the mammalian host. *Nat. Rev. Microbiol.*, **15**, 96–108.

120. Vinces,M.D. and Kumamoto,C.A. (2007) The morphogenetic regulator Czf1p is a DNA-binding protein that regulates white opaque switching in Candida albicans. *Microbiology*, **153**, 2877–2884.

121. Noffz,C.S., Liedschulte,V., Lengeler,K. and Ernst,J.F. (2008) Functional Mapping of the Candida albicans Efg1 Regulator. *Eukaryotic Cell*, **7**, 881–893.

122. Hernday,A.D., Lohse,M.B., Fordyce,P.M., Nobile,C.J., DeRisi,J.L. and Johnson,A.D. (2013) Structure of the transcriptional network controlling white-opaque switching in *C andida albicans*: White-opaque switching network in *C . albicans*. *Molecular Microbiology*, 10.1111/mmi.12329.

123. Holland,L.M., Schröder,M.S., Turner,S.A., Taff,H., Andes,D., Grózer,Z., Gácser,A., Ames,L., Haynes,K., Higgins,D.G., *et al.* (2014) Comparative Phenotypic Analysis of the Major Fungal Pathogens Candida parapsilosis and Candida albicans. *PLoS Pathogens*, **10**, e1004365.

124. Connolly,L.A., Riccombeni,A., Grózer,Z., Holland,L.M., Lynch,D.B., Andes,D.R., Gácser,A. and Butler,G. (2013) The APSES transcription factor Efg1 is a global regulator that controls morphogenesis and biofilm formation in *Candida parapsilosis*: Efg1 in *Candida parapsilosis*. *Molecular Microbiology*, 10.1111/mmi.12345.

125. Guasch,A., Aranguren-Ibáñez,Á., Pérez-Luque,R., Aparicio,D., Martínez-Høyer,S., Mulero,M.C., Serrano-Candelas,E., Pérez-Riba,M. and Fita,I. (2015) Calcineurin Undergoes a Conformational Switch Evoked via Peptidyl-Prolyl Isomerization. *PLoS ONE*, **10**, e0134569.

126. Oates,M.E., Romero,P., Ishida,T., Ghalwash,M., Mizianty,M.J., Xue,B., Dosztanyi,Z., Uversky,V.N., Obradovic,Z., Kurgan,L., *et al.* (2013) D2P2: database of disordered protein predictions. *Nucleic Acids Research*, **41**, D508–D516.

127. Ou,H.D., Phan,S., Deerinck,T.J., Thor,A., Ellisman,M.H. and O'Shea,C.C. (2017) ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*, **357**.

128. Sheinman,M., Bénichou,O., Kafri,Y. and Voituriez,R. (2012) Classes of fast and specific search mechanisms for proteins on DNA. *Rep Prog Phys*, **75**, 026601.

129. Groothuizen,F.S. and Sixma,T.K. (2016) The conserved molecular machinery in DNA mismatch repair enzyme structures. *DNA Repair*, **38**, 14–23.

130. Pluciennik,A., Burdett,V., Baitinger,C., Iyer,R.R., Shi,K. and Modrich,P. (2013) Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLα endonuclease activation. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 12277–12282.

131. Fyodorov,D.V. and Kadonaga,J.T. (2003) Chromatin Assembly In Vitro with Purified Recombinant ACF and NAP-1. In *Methods in Enzymology*. Elsevier, Vol. 371, pp. 499–515.

132. Dyer,P.N., Edayathumangalam,R.S., White,C.L., Bao,Y., Chakravarthy,S., Muthurajan,U.M. and Luger,K. (2004) Reconstitution of nucleosome core particles from recombinant histones and DNA. *Meth. Enzymol.*, **375**, 23–44.

133. Stein,A. (1989) [28] Reconstitution of chromatin from purified components. In *Methods in Enzymology*. Elsevier, Vol. 170, pp. 585–603.

134. Lusser,A. and Kadonaga,J.T. (2004) Strategies for the reconstitution of chromatin. *Nature Methods*, **1**, 19–26.

135. Studying Chromatin Dynamics In Vitro: Chromatin Assembly, Remodeling, and Transcription.

136. Arents,G., Burlingame,R.W., Wang,B.C., Love,W.E. and Moudrianakis,E.N. (1991) The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences*, **88**, 10148–10152.

137. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251–260.

138. Hansen,J.C., van Holde,K.E. and Lohr,D. (1991) The mechanism of nucleosome assembly onto oligomers of the sea urchin 5 S DNA positioning sequence. *J. Biol. Chem.*, **266**, 4276–4282.

139. Simpson,R.T. and Stafford,D.W. (1983) Structural features of a phased nucleosome core particle. *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 51–55.

140. Lowary,P.. and Widom,J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*, **276**, 19–42.

141. Shundrovsky,A., Smith,C.L., Lis,J.T., Peterson,C.L. and Wang,M.D. (2006) Probing SWI/SNF remodeling of the nucleosome by unzipping single DNA molecules. *Nature Structural & Molecular Biology*, **13**, 549–554.

142. Visnapuu,M.-L., Fazio,T., Wind,S. and Greene,E.C. (2008) Parallel arrays of geometric nanowells for assembling curtains of DNA with controlled lateral dispersion. *Langmuir*, **24**, 11293–11299.

143. Vogler,C., Huber,C., Waldmann,T., Ettig,R., Braun,L., Izzo,A., Daujat,S., Chassignet,I., Lopez-Contreras,A.J., Fernandez-Capetillo,O., *et al.* (2010) Histone H2A C-Terminus Regulates Chromatin Dynamics, Remodeling, and Histone H1 Binding. *PLoS Genetics*, **6**, e1001234.

144. Fink,A.L. (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des*, **3**, R9-23.

145. Gopal,G.J. and Kumar,A. (2013) Strategies for the Production of Recombinant Protein in Escherichia coli. *The Protein Journal*, **32**, 419–425.

146. Hwang,P.M., Pan,J.S. and Sykes,B.D. (2014) Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in Escherichia coli. *FEBS Letters*, **588**, 247–252.

147. Klinker,H., Haas,C., Harrer,N., Becker,P.B. and Mueller-Planitz,F. (2014) Rapid Purification of Recombinant Histones. *PLoS ONE*, **9**, e104029.

148. Luger,K. (2003) Structure and dynamic behavior of nucleosomes. *Curr. Opin. Genet. Dev.*, **13**, 127–135.

149. Arimura,Y., Tachiwana,H., Oda,T., Sato,M. and Kurumizaka,H. (2012) Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome. *Biochemistry*, **51**, 3302–3309.

150. Irvine,D., Tuerk,C. and Gold,L. (1991) Selexion. *Journal of Molecular Biology*, **222**, 739–761.

151. Vasudevan,D., Chua,E.Y.D. and Davey,C.A. (2010) Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol.*, **403**, 1–10.

152. Chua,E.Y.D., Vasudevan,D., Davey,G.E., Wu,B. and Davey,C.A. (2012) The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res.*, **40**, 6338–6352.

153. Cui,F. and Zhurkin,V.B. (2010) Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro. *J. Biomol. Struct. Dyn.*, **27**, 821–841.

154. Burchell,L., Chaugule,V.K. and Walden,H. (2012) Small, N-Terminal Tags Activate Parkin E3 Ubiquitin Ligase Activity by Disrupting Its Autoinhibited Conformation. *PLoS ONE*, **7**, e34748.

155. Chan,W.C.W., Maxwell,D.J., Gao,X., Bailey,R.E., Han,M. and Nie,S. (2002) Luminescent quantum dots for multiplexed biological detection and imaging. *Curr. Opin. Biotechnol.*, **13**, 40–46.

156. Wegner,K.D. and Hildebrandt,N. (2015) Quantum dots: bright and versatile in vitro and in vivo fluorescence imaging biosensors. *Chem. Soc. Rev.*, 10.1039/C4CS00532E.

157. Hohng,S. and Ha,T. (2004) Near-Complete Suppression of Quantum Dot Blinking in Ambient Conditions. *Journal of the American Chemical Society*, **126**, 1324–1325.

158. Dunker,A.K. (1975) Analysis of protein-nucleic acid filter binding using the Poisson distribution: a method to estimate co-operativity. *FEBS Lett.*, **52**, 323–326.

159. Simpson,R.T., Thoma,F. and Brubaker,J.M. (1985) Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: a model system for study of higher order structure. *Cell*, **42**, 799–808.

160. Hansen,J.C. and Lohr,D. (1993) Assembly and structural properties of subsaturated chromatin arrays. *J. Biol. Chem.*, **268**, 5840–5848.

161. Tillo,D. and Hughes,T.R. (2009) G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics*, **10**, 442.

162. Visnapuu,M.-L. and Greene,E.C. (2009) Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. *Nat. Struct. Mol. Biol.*, **16**, 1056–1062.

163. Gorman,J., Plys,A.J., Visnapuu,M.-L., Alani,E. and Greene,E.C. (2010) Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.*, **17**, 932–938.

164. Rogacheva,M.V., Manhart,C.M., Chen,C., Guarne,A., Surtees,J. and Alani,E. (2014) Mlh1-Mlh3, a Meiotic Crossover and DNA Mismatch Repair Factor, Is a Msh2-Msh3-stimulated Endonuclease. *Journal of Biological Chemistry*, **289**, 5664–5673.

165. Adkins,N.L., Niu,H., Sung,P. and Peterson,C.L. (2013) Nucleosome dynamics regulates DNA processing. *Nature Structural & Molecular Biology*, **20**, 836–842.

166. Chen,X., Cui,D., Papusha,A., Zhang,X., Chu,C.-D., Tang,J., Chen,K., Pan,X. and Ira,G. (2012) The Fun30 nucleosome remodeller promotes resection of DNA double-strand break ends. *Nature*, **489**, 576–580.

167. Oldfield,C.J. and Dunker,A.K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.*, **83**, 553–584.

168. van Leeuwen,H.C., Strating,M.J., Rensen,M., de Laat,W. and van der Vliet,P.C. (1997) Linker length and composition influence the flexibility of Oct-1 DNA binding. *EMBO J.*, **16**, 2043–2053.

169. Currie,S.L., Lau,D.K.W., Doane,J.J., Whitby,F.G., Okon,M., McIntosh,L.P. and Graves,B.J. (2017) Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4 and ETV5. *Nucleic Acids Res.*, **45**, 2223–2241.

170. Kozlov,A.G., Shinn,M.K., Weiland,E.A. and Lohman,T.M. (2017) Glutamate promotes SSB protein-protein Interactions via intrinsically disordered regions. *J. Mol. Biol.*, **429**, 2790–2801.

171. Vuzman,D. and Levy,Y. (2012) Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol. Biosyst.*, **8**, 47–57.

172. Bronner,C.E., Baker,S.M., Morrison,P.T., Warren,G., Smith,L.G., Lescoe,M.K., Kane,M., Earabino,C., Lipford,J., Lindblom,A., *et al.* (1994) Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer. *Nature*, **368**, 258–261.

173. Prolla,T.A., Christie,D.M. and Liskay,R.M. (1994) Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs of the bacterial mutL gene. *Mol. Cell. Biol.*, **14**, 407–415.

174. Hall,M.C., Shcherbakova,P.V., Fortune,J.M., Borchers,C.H., Dial,J.M., Tomer,K.B. and Kunkel,T.A. (2003) DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair. *Nucleic Acids Res.*, **31**, 2025–2034.

175. Argueso,J.L., Kijas,A.W., Sarin,S., Heck,J., Waase,M. and Alani,E. (2003) Systematic mutagenesis of the Saccharomyces cerevisiae MLH1 gene reveals distinct roles for Mlh1p in meiotic crossing over and in vegetative and meiotic mismatch repair. *Mol. Cell. Biol.*, **23**, 873–886.

176. Groothuizen,F.S., Winkler,I., Cristóvão,M., Fish,A., Winterwerp,H.H., Reumer,A., Marx,A.D., Hermans,N., Nicholls,R.A., Murshudov,G.N., *et al.* (2015) MutS/MutL

crystal structure reveals that the MutS sliding clamp loads MutL onto DNA. *eLife*, **4**.

177. Kunkel,T.A. and Erie,D.A. (2015) Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu. Rev. Genet.*, **49**, 291–313.

178. Kadyrov,F.A., Dzantiev,L., Constantin,N. and Modrich,P. (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*, **126**, 297–308.

179. Kadyrov,F.A., Genschel,J., Fang,Y., Penland,E., Edelmann,W. and Modrich,P. (2009) A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 8495–8500.

180. Genschel,J., Kadyrova,L.Y., Iyer,R.R., Dahal,B.K., Kadyrov,F.A. and Modrich,P. (2017) Interaction of proliferating cell nuclear antigen with PMS2 is required for MutLα activation and function in mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 4930–4935.

181. Lu,A.L. (1987) Influence of GATC sequences on Escherichia coli DNA mismatch repair in vitro. *J. Bacteriol.*, **169**, 1254–1259.

182. Ou,J., Niessen,R.C., Vonk,J., Westers,H., Hofstra,R.M.W. and Sijmons,R.H. (2008) A database to support the interpretation of human mismatch repair gene variants. *Hum. Mutat.*, **29**, 1337–1341.

183. Claeys Bouuaert,C. and Keeney,S. (2017) Distinct DNA-binding surfaces in the ATPase and linker domains of MutLγ determine its substrate specificities and exert separable functions in meiotic recombination and mismatch repair. *PLoS Genet.*, **13**, e1006722.

184. Räschle,M., Dufner,P., Marra,G. and Jiricny,J. (2002) Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha. *J. Biol. Chem.*, **277**, 21810–21820.

185. Hall,M.C. and Kunkel,T.A. (2001) Purification of eukaryotic MutL homologs from Saccharomyces cerevisiae using self-affinity technology. *Protein Expr. Purif.*, **21**, 333–342.

186. Cooper,M.P., Balajee,A.S. and Bohr,V.A. (1999) The C-terminal domain of p21 inhibits nucleotide excision repair In vitro and In vivo. *Mol. Biol. Cell*, **10**, 2119–2129.

187. Finkelstein,J., Antony,E., Hingorani,M.M. and O'Donnell,M. (2003) Overproduction and analysis of eukaryotic multiprotein complexes in Escherichia coli using a dual-vector strategy. *Anal. Biochem.*, **319**, 78–87.

188. Modesti,M. (2011) Fluorescent labeling of proteins. *Methods Mol. Biol.*, **783**, 101–120.

189. Rogacheva,M.V., Manhart,C.M., Chen,C., Guarne,A., Surtees,J. and Alani,E. (2014) Mlh1-Mlh3, a meiotic crossover and DNA mismatch repair factor, is a Msh2-Msh3-stimulated endonuclease. *J. Biol. Chem.*, **289**, 5664–5673.

190. Manhart,C.M., Ni,X., White,M.A., Ortega,J., Surtees,J.A. and Alani,E. (2017) The mismatch repair and meiotic recombination endonuclease Mlh1-Mlh3 is activated by polymer formation and can cleave DNA substrates in trans. *PLoS Biol.*, **15**, e2001164.

191. Alani,E., Chi,N.W. and Kolodner,R. (1995) The Saccharomyces cerevisiae Msh2 protein specifically binds to duplex oligonucleotides containing mismatched DNA base pairs and insertions. *Genes Dev.*, **9**, 234–247.

192. Rose MD, Winston F and Hieter P (1990) Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual Cold Spring Harbor Laboratory Press.

193. Heck,J.A., Argueso,J.L., Gemici,Z., Reeves,R.G., Bernard,A., Aquadro,C.F. and Alani,E. (2006) Negative epistasis between natural variants of the Saccharomyces cerevisiae MLH1 and PMS1 genes results in a defect in mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 3256–3261.

194. Gietz,R.D., Schiestl,R.H., Willems,A.R. and Woods,R.A. (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, **11**, 355–360.

195. Christianson,T.W., Sikorski,R.S., Dante,M., Shero,J.H. and Hieter,P. (1992) Multifunctional yeast high-copy-number shuttle vectors. *Gene*, **110**, 119–122.

196. Tran,H.T., Keen,J.D., Kricker,M., Resnick,M.A. and Gordenin,D.A. (1997) Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol. Cell. Biol.*, **17**, 2859–2865.

197. Hammer,Ø., Harper,D.A. and Ryan,P.D. (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, **4**, 1–9.

198. Efron,B. and Tibshirani,R.J. (1993) An Introduction to the Bootstrap Softcover reprint of the original 1st ed. 1993 edition. Chapman and Hall/CRC, New York.

199. Kim,Y., de la Torre,A., Leal,A.A. and Finkelstein,I.J. (2017) Efficient modification of λ-DNA substrates for single-molecule studies. *Sci. Rep.*, **7**, 2071.

200. Yu,H., Singh Gautam,A.K., Wilmington,S.R., Wylie,D., Martinez-Fonts,K., Kago,G., Warburton,M., Chavali,S., Inobe,T., Finkelstein,I.J., *et al.* (2016) Conserved sequence preferences contribute to substrate recognition by the proteasome. *J. Biol. Chem.*, **291**, 14526–14539.

201. Fishbain,S., Inobe,T., Israeli,E., Chavali,S., Yu,H., Kago,G., Babu,M.M. and Matouschek,A. (2015) Sequence composition of disordered regions fine-tunes protein half-life. *Nat. Struct. Mol. Biol.*, **22**, 214–221.

202. Pluciennik,A., Dzantiev,L., Iyer,R.R., Constantin,N., Kadyrov,F.A. and Modrich,P. (2010) PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 16066–16071.

203. Kawasoe,Y., Tsurimoto,T., Nakagawa,T., Masukata,H. and Takahashi,T.S. (2016) MutSα maintains the mismatch repair capability by inhibiting PCNA unloading. *eLife*, **5**.

204. Liu,J., Hanne,J., Britton,B.M., Bennett,J., Kim,D., Lee,J.-B. and Fishel,R. (2016) Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature*, **539**, 583–587.

205. Gallardo,I.F., Pasupathy,P., Brown,M., Manhart,C.M., Neikirk,D.P., Alani,E. and Finkelstein,I.J. (2015) High-throughput universal DNA curtain arrays for single-molecule fluorescence imaging. *Langmuir*, **31**, 10310–10317.

206. Soniat,M.M., Myler,L.R., Schaub,J.M., Kim,Y., Gallardo,I.F. and Finkelstein,I.J. (2017) Next-Generation DNA Curtains for Single-Molecule Studies of Homologous Recombination. *Meth. Enzymol.*, **592**, 259–281.

207. Smith,C.E., Bowen,N., Graham,W.J., Goellner,E.M., Srivatsan,A. and Kolodner,R.D. (2015) Activation of Saccharomyces cerevisiae Mlh1-Pms1 Endonuclease in a Reconstituted Mismatch Repair System. *J. Biol. Chem.*, **290**, 21580–21590.

208. Sakato,M., Zhou,Y. and Hingorani,M.M. (2012) ATP binding and hydrolysis-driven rate-determining events in the RFC-catalyzed PCNA clamp loading reaction. *J. Mol. Biol.*, **416**, 176–191.

209. Dionne,I., Brown,N.J., Woodgate,R. and Bell,S.D. (2008) On the mechanism of loading the PCNA sliding clamp by RFC. *Mol. Microbiol.*, **68**, 216–222.

210. Pillon,M.C., Babu,V.M.P., Randall,J.R., Cai,J., Simmons,L.A., Sutton,M.D. and Guarné,A. (2015) The sliding clamp tethers the endonuclease domain of MutL to DNA. *Nucleic Acids Res.*, **43**, 10746–10759.

211. Lao,Y., Lee,C.G. and Wold,M.S. (1999) Replication protein A interactions with DNA. 2. Characterization of double-stranded DNA-binding/helix-destabilization activities and the role of the zinc-finger domain in DNA interactions. *Biochemistry*, **38**, 3974–3984.

212. Kim,C., Snyder,R.O. and Wold,M.S. (1992) Binding properties of replication protein A from human and yeast cells. *Mol. Cell. Biol.*, **12**, 3050–3059.

213. Blackwell,L.J. and Borowiec,J.A. (1994) Human replication protein A binds single-stranded DNA in two distinct complexes. *Mol. Cell. Biol.*, **14**, 3993–4001.

214. Obradovic,Z., Peng,K., Vucetic,S., Radivojac,P., Brown,C.J. and Dunker,A.K. (2003) Predicting intrinsic disorder from amino acid sequence. *Proteins*, **53 Suppl 6**, 566–572.

215. Finkelstein,I.J. and Greene,E.C. (2013) Molecular traffic jams on DNA. *Annu Rev Biophys*, **42**, 241–263.

216. Kemmerich,F.E., Swoboda,M., Kauert,D.J., Grieb,M.S., Hahn,S., Schwarz,F.W., Seidel,R. and Schlierf,M. (2016) Simultaneous Single-Molecule Force and Fluorescence Sampling of DNA Nanostructure Conformations Using Magnetic Tweezers. *Nano Lett.*, **16**, 381–386.

217. Liu,S., Chistol,G., Hetherington,C.L., Tafoya,S., Aathavan,K., Schnitzbauer,J., Grimes,S., Jardine,P.J. and Bustamante,C. (2014) A Viral Packaging Motor Varies Its DNA Rotation and Step Size to Preserve Subunit Coordination as the Capsid Fills. *Cell*, **157**, 702–713.

218. Thorpe,H.M. and Smith,M.C. (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 5505–5510.

219. Bell,J.C., Plank,J.L., Dombrowski,C.C. and Kowalczykowski,S.C. (2012) Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. *Nature*, **491**, 274–8.

220. Luzzietti,N., Knappe,S., Richter,I. and Seidel,R. (2012) Nicking enzyme-based internal labeling of DNA at multiple loci. *Nat. Protocols*, **7**, 643–653.

221. Luzzietti,N., Brutzer,H., Klaue,D., Schwarz,F.W., Staroske,W., Clausing,S. and Seidel,R. (2011) Efficient preparation of internally modified single-molecule constructs using nicking enzymes. *Nucl. Acids Res.*, **39**, e15.

222. Berezney,J.P. and Saleh,O.A. (2014) Locked nucleic acid oligomers as handles for single molecule manipulation. *Nucl. Acids Res.*, **42**, e150.

223. Zohar,H., Hetherington,C.L., Bustamante,C. and Muller,S.J. (2010) Peptide Nucleic Acids as Tools for Single-Molecule Sequence Detection and Manipulation. *Nano Lett.*, **10**, 4697–4701.

224. Zohar,H. and Muller,S.J. (2011) Labeling DNA for single-molecule experiments: methods of labeling internal specific sequences on double-stranded DNA. *Nanoscale*, **3**, 3027–3039.

225. Pluciennik,A., Burdett,V., Baitinger,C., Iyer,R.R., Shi,K. and Modrich,P. (2013) Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLα endonuclease activation. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 12277–12282.

226. Iyer,R.R., Pluciennik,A., Napierala,M. and Wells,R.D. (2015) DNA triplet repeat expansion and mismatch repair. *Annual Review of Biochemistry*, **84**, 199–226.

227. Schmidt,M.H.M. and Pearson,C.E. (2016) Disease-associated repeat instability and mismatch repair. *DNA Repair (Amst.)*, **38**, 117–126.

228. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 6640–6645.

229. Sambrook,J. and Russell,D.W. (2006) Transformation of E. coli by Electroporation. *Cold Spring Harb Protoc*, **2006**, pdb.prot3933.

230. De Biasio,A., Sánchez,R., Prieto,J., Villate,M., Campos-Olivas,R. and Blanco,F.J. (2011) Reduced Stability and Increased Dynamics in the Human Proliferating Cell Nuclear Antigen (PCNA) Relative to the Yeast Homolog. *PLoS ONE*, **6**, e16600.

231. Finkelstein,J., Antony,E., Hingorani,M.M. and O'Donnell,M. (2003) Overproduction and analysis of eukaryotic multiprotein complexes in Escherichia coli using a dual-vector strategy. *Anal. Biochem.*, **319**, 78–87.

232. Gomes,X.V. (2001) ATP Utilization by Yeast Replication Factor C. II. MULTIPLE STEPWISE ATP BINDING EVENTS ARE REQUIRED TO LOAD PROLIFERATING CELL NUCLEAR ANTIGEN ONTO PRIMED DNA. *Journal of Biological Chemistry*, **276**, 34776–34783.

233. Dieckman,L.M., Freudenthal,B.D. and Washington,M.T. (2012) PCNA Structure and Function: Insights from Structures of PCNA Complexes and Post-translationally Modified PCNA. In MacNeill,S. (ed), *The Eukaryotic Replisome: a Guide to Protein Structure and Function*. Springer Netherlands, Dordrecht, Vol. 62, pp. 281–299.

234. Krishna,T.S., Kong,X.P., Gary,S., Burgers,P.M. and Kuriyan,J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*, **79**, 1233–1243.

235. Kochaniak,A.B., Habuchi,S., Loparo,J.J., Chang,D.J., Cimprich,K.A., Walter,J.C. and van Oijen,A.M. (2009) Proliferating cell nuclear antigen uses two distinct modes to move along DNA. *J. Biol. Chem.*, **284**, 17700–17710.

236. Yao,N., Hurwitz,J. and O'Donnell,M. (2000) Dynamics of beta and proliferating cell nuclear antigen sliding clamps in traversing DNA secondary structure. *J. Biol. Chem.*, **275**, 1421–1432.

237. Biyani,M. and Nishigaki,K. (2005) Single-Strand Conformation Polymorphism (SSCP) of Oligodeoxyribonucleotides: An Insight into Solution Structural Dynamics of DNAs Provided by Gel Electrophoresis and Molecular Dynamics Simulations. *J Biochem*, **138**, 363–373.

238. Mariappan,S.V., Silks,L.A., Chen,X., Springer,P.A., Wu,R., Moyzis,R.K., Bradbury,E.M., Garcia,A.E. and Gupta,G. (1998) Solution structures of the Huntington's disease DNA triplets, (CAG)n. *J. Biomol. Struct. Dyn.*, **15**, 723–744.

239. Amrane,S., Saccà,B., Mills,M., Chauhan,M., Klump,H.H. and Mergny,J.-L. (2005) Length-dependent energetics of (CTG)n and (CAG)n trinucleotide repeats. *Nucl. Acids Res.*, **33**, 4065–4077.

240. Lee,B.J., Barch,M., Castner,Edward W., Völker,J. and Breslauer,K.J. (2007) Structure and Dynamics in DNA Looped Domains: CAG Triplet Repeat Sequence Dynamics Probed by 2-Aminopurine Fluorescence. *Biochemistry*, **46**, 10756–10766.

241. Degtyareva,N.N., Barber,C.A., Reddish,M.J. and Petty,J.T. (2011) Sequence Length Dictates Repeated CAG Folding in Three-Way Junctions. *Biochemistry*, **50**, 458–465.

242. Efron, Bradley and Tibshirani, R.J. (1994) An Introduction to the Bootstrap CRC Press.

243. Kochaniak,A.B., Habuchi,S., Loparo,J.J., Chang,D.J., Cimprich,K.A., Walter,J.C. and van Oijen,A.M. (2009) Proliferating cell nuclear antigen uses two distinct modes to move along DNA. *The Journal of biological chemistry*, **284**, 17700–10.

244. March,Z.M., King,O.D. and Shorter,J. (2016) Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Research*, **1647**, 9–18.

245. Lancaster,A.K., Nutter-Upham,A., Lindquist,S. and King,O.D. (2014) PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics*, **30**, 2501–2502.

246. Fowler,D.M. and Kelly,J.W. (2012) Functional amyloidogenesis and cytotoxicity-insights into biology and pathology. *PLoS Biol.*, **10**, e1001459.

247. Sudhakaran,I.P. and Ramaswami,M. (2017) Long-term memory consolidation: The role of RNA-binding proteins with prion-like domains. *RNA Biology*, **14**, 568–586.

248. King,O.D., Gitler,A.D. and Shorter,J. (2012) The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res.*, **1462**, 61–80.

249. Moorthy,S.D., Davidson,S., Shchuka,V.M., Singh,G., Malek-Gilani,N., Langroudi,L., Martchenko,A., So,V., Macpherson,N.N. and Mitchell,J.A. (2017) Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Research*, **27**, 246–258.

250. Cho,W.-K., Spille,J.-H., Hecht,M., Lee,C., Li,C., Grube,V. and Cisse,I.I. (2018) Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*, **361**, 412–415.

251. Si,H., Hernday,A.D., Hirakawa,M.P., Johnson,A.D. and Bennett,R.J. (2013) Candida albicans white and opaque cells undergo distinct programs of filamentous growth. *PLoS Pathog.*, **9**, e1003210.

252. Lohse,M.B., Zordan,R.E., Cain,C.W. and Johnson,A.D. (2010) Distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in Candida albicans. *Proceedings of the National Academy of Sciences*, **107**, 14105–14110.

253. Lassak,T., Schneider,E., Bussmann,M., Kurtz,D., Manak,J.R., Srikantha,T., Soll,D.R. and Ernst,J.F. (2011) Target specificity of the Candida albicans Efg1 regulator: Efg1 target recognition. *Molecular Microbiology*, **82**, 602–618.

254. Gallardo,I.F., Pasupathy,P., Brown,M., Manhart,C.M., Neikirk,D.P., Alani,E. and Finkelstein,I.J. (2015) High-Throughput Universal DNA Curtain Arrays for Single-Molecule Fluorescence Imaging. *Langmuir*, **31**, 10310–10317.

255. Brown,M.W., Kim,Y., Williams,G.M., Huck,J.D., Surtees,J.A. and Finkelstein,I.J. (2016) Dynamic DNA binding licenses a repair factor to bypass roadblocks in search of DNA lesions. *Nat Commun*, **7**, 10607.

256. Myler,L.R., Gallardo,I.F., Soniat,M.M., Deshpande,R.A., Gonzalez,X.B., Kim,Y., Paull,T.T. and Finkelstein,I.J. (2017) Single-Molecule Imaging Reveals How Mre11-Rad50-Nbs1 Initiates DNA Break Repair. *Molecular Cell*, **67**, 891-898.e4.

257. Hnisz,D. and Young,R.A. (2017) New Insights into Genome Structure: Genes of a Feather Stick Together. *Molecular Cell*, **67**, 730–731.

258. Kwon,I., Kato,M., Xiang,S., Wu,L., Theodoropoulos,P., Mirzaei,H., Han,T., Xie,S., Corden,J.L. and McKnight,S.L. (2013) Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell*, **155**, 1049–1060.

259. Harlen,K.M. and Churchman,L.S. (2017) The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology*, **18**, 263–273.

260. Sekinger,E.A., Moqtaderi,Z. and Struhl,K. (2005) Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell*, **18**, 735–748.

261. Bintu,L., Ishibashi,T., Dangkulwanich,M., Wu,Y.-Y., Lubkowska,L., Kashlev,M. and Bustamante,C. (2012) Nucleosomal elements that control the topography of the barrier to transcription. *Cell*, **151**, 738–749.

262. Bai,L. and Morozov,A.V. (2010) Gene regulation by nucleosome positioning. *Trends Genet.*, **26**, 476–483.

263. Nobile,C.J., Fox,E.P., Nett,J.E., Sorrells,T.R., Mitrovich,Q.M., Hernday,A.D., Tuch,B.B., Andes,D.R. and Johnson,A.D. (2012) A Recently Evolved Transcriptional Network Controls Biofilm Development in Candida albicans. *Cell*, **148**, 126–138.

264. Yano,J., Yu,A., Fidel,P.L. and Noverr,M.C. (2016) Transcription Factors Efg1 and Bcr1 Regulate Biofilm Formation and Virulence during Candida albicans-Associated Denture Stomatitis. *PLOS ONE*, **11**, e0159692.

265. Prilusky,J., Felder,C.E., Zeev-Ben-Mordehai,T., Rydberg,E.H., Man,O., Beckmann,J.S., Silman,I. and Sussman,J.L. (2005) FoldIndex(C): a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, **21**, 3435–3438.

266. Gueneau,E., Dherin,C., Legrand,P., Tellier-Lebegue,C., Gilquin,B., Bonnesoeur,P., Londino,F., Quemener,C., Le Du,M.-H., Márquez,J.A., *et al.* (2013) Structure of the MutLα C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. *Nature Structural & Molecular Biology*, **20**, 461–468.

267. Dherin,C., Gueneau,E., Francin,M., Nunez,M., Miron,S., Liberti,S.E., Rasmussen,L.J., Zinn-Justin,S., Gilquin,B., Charbonnier,J.-B., *et al.* (2009) Characterization of a Highly Conserved Binding Site of Mlh1 Required for Exonuclease I-Dependent Mismatch Repair. *Molecular and Cellular Biology*, **29**, 907–918.

268. Genschel,J., Kadyrova,L.Y., Iyer,R.R., Dahal,B.K., Kadyrov,F.A. and Modrich,P. (2017) Interaction of proliferating cell nuclear antigen with PMS2 is required for MutLα activation and function in mismatch repair. *Proceedings of the National Academy of Sciences*, 10.1073/pnas.1702561114.

269. Ando,T. and Skolnick,J. (2014) Sliding of proteins non-specifically bound to DNA: Brownian dynamics studies with coarse-grained protein and DNA models. *PLoS Comput. Biol.*, **10**, e1003990.

270. Guarné,A., Ramon-Maiques,S., Wolff,E.M., Ghirlando,R., Hu,X., Miller,J.H. and Yang,W. (2004) Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *The EMBO Journal*, **23**, 4134–4145.

271. Claeys Bouuaert,C. and Keeney,S. (2017) Distinct DNA-binding surfaces in the ATPase and linker domains of MutLγ determine its substrate specificities and exert separable functions in meiotic recombination and mismatch repair. *PLOS Genetics*, **13**, e1006722.

272. Kadyrova,L.Y., Blanko,E.R. and Kadyrov,F.A. (2011) CAF-I-dependent control of degradation of the discontinuous strands during mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 2753–2758.

273. Kadyrova,L.Y., Mertz,T.M., Zhang,Y., Northam,M.R., Sheng,Z., Lobachev,K.S., Shcherbakova,P.V. and Kadyrov,F.A. (2013) A reversible histone H3 acetylation

cooperates with mismatch repair and replicative polymerases in maintaining genome stability. *PLoS Genet.*, **9**, e1003899.

274. Laguri,C., Duband-Goulet,I., Friedrich,N., Axt,M., Belin,P., Callebaut,I., Gilquin,B., Zinn-Justin,S. and Couprie,J. (2008) Human mismatch repair protein MSH6 contains a PWWP domain that targets double stranded DNA. *Biochemistry*, **47**, 6199–6207.

275. Qin,S. and Min,J. (2014) Structure and function of the nucleosome-binding PWWP domain. *Trends Biochem. Sci.*, **39**, 536–547.

276. Lee,S.D. and Alani,E. (2006) Analysis of Interactions Between Mismatch Repair Initiation Factors and the Replication Processivity Factor PCNA. *Journal of Molecular Biology*, **355**, 175–184.

277. Gorman,J., Wang,F., Redding,S., Plys,A.J., Fazio,T., Wind,S., Alani,E.E. and Greene,E.C. (2012) Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E3074-3083.

278. Clark,A.B., Valle,F., Drotschmann,K., Gary,R.K. and Kunkel,T.A. (2000) Functional interaction of proliferating cell nuclear antigen with MSH2-MSH6 and MSH2-MSH3 complexes. *J. Biol. Chem.*, **275**, 36498–36501.

279. Pluciennik,A., Dzantiev,L., Iyer,R.R., Constantin,N., Kadyrov,F.A. and Modrich,P. (2010) PCNA function in the activation and strand direction of MutL endonuclease in mismatch repair. *Proceedings of the National Academy of Sciences*, **107**, 16066–16071.

280. Kawasoe,Y., Tsurimoto,T., Nakagawa,T., Masukata,H. and Takahashi,T.S. (2016) MutSα maintains the mismatch repair capability by inhibiting PCNA unloading. *eLife*, **5**, e15155.

281. Iyer,R.R., Pluciennik,A., Genschel,J., Tsai,M.-S., Beese,L.S. and Modrich,P. (2010) MutLalpha and proliferating cell nuclear antigen share binding sites on MutSbeta. *J. Biol. Chem.*, **285**, 11730–11739.

282. Gacy,A.M., Goellner,G., Juranić,N., Macura,S. and McMurray,C.T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell*, **81**, 533–540.

283. Viterbo,D., Michoud,G., Mosbach,V., Dujon,B. and Richard,G.-F. (2016) Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. *DNA Repair*, **42**, 94–106.

284. Chahwan,R., Edelmann,W., Scharff,M.D. and Roa,S. (2011) Mismatch-mediated error prone repair at the immunoglobulin genes. *Biomedicine & Pharmacotherapy*, **65**, 529–536.

285. Eccleston,J., Yan,C., Yuan,K., Alt,F.W. and Selsing,E. (2011) Mismatch Repair Proteins MSH2, MLH1, and EXO1 Are Important for Class-Switch Recombination Events Occurring in B Cells That Lack Nonhomologous End Joining. *The Journal of Immunology*, **186**, 2336–2343.

286. Chahwan,R., van Oers,J.M.M., Avdievich,E., Zhao,C., Edelmann,W., Scharff,M.D. and Roa,S. (2012) The ATPase activity of MLH1 is required to orchestrate DNA double-strand breaks and end processing during class switch recombination. *The Journal of Experimental Medicine*, **209**, 671–678.

287. Martin,A., Li,Z., Lin,D.P., Bardwell,P.D., Iglesias-Ussel,M.D., Edelmann,W. and Scharff,M.D. (2003) Msh2 ATPase activity is essential for somatic hypermutation at a-T basepairs and for efficient class switch recombination. *J. Exp. Med.*, **198**, 1171–1178.

288. Stavnezer,J., Björkman,A., Du,L., Cagigi,A. and Pan-Hammarström,Q. (2010) Mapping of switch recombination junctions, a tool for studying DNA repair pathways during immunoglobulin class switching. *Adv. Immunol.*, **108**, 45–109.

289. Larson,A.G., Elnatan,D., Keenen,M.M., Trnka,M.J., Johnston,J.B., Burlingame,A.L., Agard,D.A., Redding,S. and Narlikar,G.J. (2017) Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*, **547**, 236–240.

290. Skalska,L., Stojnic,R., Li,J., Fischer,B., Cerda-Moya,G., Sakai,H., Tajbakhsh,S., Russell,S., Adryan,B. and Bray,S.J. (2015) Chromatin signatures at Notch-regulated enhancers reveal large-scale changes in H3K56ac upon activation. *EMBO J.*, **34**, 1889–1904.

291. Kim,J., Park,S. and Lee,J.-S. (2018) Epigenetic Control of Oxidative Stresses by Histone Acetyltransferases in Candida albicans. *J. Microbiol. Biotechnol.*, **28**, 181–189.

Vita

Yoori Kim was born in Seoul, South Korea. After completing her work at Soong-eui High School, Seoul, South Korea, in 2004, she entered Sogang University in Seoul, South Korea. She received the degree of Bachelor of Science from Sogang University in February 2009. During following two years, she was a graduate student at Sogang University, and received the degree of Master of Science from Sogang University in February 2011. She was employed as a research scientist at KIST, Seoul, South Korea in 2011-2012. In September, 2013, she entered the Graduate School at the University of Texas at Austin.

Address: danielyoori@gmail.com

This manuscript was typed by the author.